Epitopes, Aggregation and Membrane Binding: Investigating the Protein Structure-Function Relationship

Author: Craig Robert Gregor

THESIS SUBMITTED IN FULFILMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

THE SCHOOL OF PHYSICS AND ASTRONOMY

THE UNIVERSITY OF EDINBURGH

2011

Acknowledgements

I am indebted to a number of people, and without whose help, completion of this work would not have been possible. I would therefore sincerely like to thank the following people for their contributions in either providing inspiration, guidance, or simply support over the course of my PhD:

Firstly I would like to thank my supervisors *Jason Crain* and *Will Hossack* for their expert guidance and motivation throughout. Secondly, I would like to thank *Eleonora Cerasoli* and *Maxim Ryadnov* for providing me with their experimental knowledge and supervision whenever I decided to venture into a laboratory. *Paul Tulip* for providing daily supervision (whether he wanted to or not) as I attempted to build upon legacy simulations. *Glenn Martyna* for whenever I had questions regarding the simulation software (provided I was prepared for the answers to be in the form of a haiku). *Raph Troitzsch* for showing me that even someone who can't count can successfully navigate a PhD. *Iain Robinson*, *Colin McGuinness*, *Chrissy Wong*, *Mira Nishimura* and *Jay Gillam* for being gracious enough to share an office and put up with me. *Tony Hart* and *Morphy Richards* for providing an outstandingly named coffee machine.

Finally I would like to thank my parents and brother, for their never ending patience, kindness and support.

Declaration

I hereby declare that the writing of this thesis is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgement has been made in the text.

signature: _____

name: _____

date: _____

Abstract

The three-dimensional structure of a protein, formed as a result of amino-acid sequences folding into compact domains, is regarded as a key factor in its biological function. How and why proteins fold into specific topologies, remain the key focus of scientific research in the field of biophysics. By stripping down complex reactions down to the most basic elements, biophysicists aim to develop simplified models for biological phenomena such as antibody discrimination, viral fusion or self-assembly.

Focusing on small model peptide systems, rather than the full proteins from which they were derived, was hoped to result in accurate structural measurements and provide a more transparent comparison between simulation and experiment. The aim of this research was therefore to investigate how accurate these models were when compared against experiment. Furthermore, while breaking down the complex biological phenomena into simple models, there was also a conscious effort to ensure that the models were representative of real biological systems, and a major focus was therefore aimed at determining whether any meaningful biomedical insight may be extrapolated from such models.

Peptides found in hormones (human chorionic gonadotropin, luteinizing hormone), viruses (HIV) and amyloid diseases (transthyretin) were selected in order to probe a variety of questions in relation to the aforementioned biological phenomena. Namely, how the primary sequence influenced the three-dimensional structure (and thus its biological function), how its environment could influence such a confirmation, and how these systems aggregated.

This doctoral study has made use of a combination of computer simulations and experimental techniques to investigate a selection of biologically relevant peptides; utilising classical atomistic molecular dynamics (MD) simulations to characterise the free-energy landscapes of the chosen peptides, and compare these findings with the secondary structure content predicted by spectroscopic methods such as circular dichroism and infrared spectroscopy. The peptide systems studied within, were found to be characterised by rugged free-energy landscapes unlike their protein counterparts (defined by singular, deep minima). Furthermore, these landscapes were found to be highly plastic and sensitive to changes in the local environment.

Contents

List of Figures

List of Tables

If we can hit that bulls-eye, the rest of the dominoes will fall like a house of cards. Checkmate!

1 Introduction

1.1 Motivation

A protein's three-dimensional structure is regarded as being a key factor in defining the functional properties of the biological role that the protein has evolved to fill. This structure arises due to amino-acid sequences in proteins folding into compact domains of specific topology (*Appendix A*). It is these domains that can act as binding sites for other molecules, or act as building blocks for larger more complex biomolecules such as viruses. One would therefore wish to predict a protein's three-dimensional structure from its constituent amino-acids, in order to shed light on the protein's biological function. However, such a prediction has been found to be fraught with difficulty and therefore, the questions of how and why a protein folds into a specific topology has remained a key focus of continuous scientific research. Such research is at the heart of the field of biophysics.

Molecular biophysics is an interdisciplinary are of research crossing the boundaries of physics, biology, chemistry and mathematics. Its purpose is to study the biophysical processes of complex macromolecular systems by focussing on intermolecular and intramolecular interactions and the resulting structural and electronic changes. The mechanisms by which proteins fold into their single native structures (corresponding to a global minimum of free energy) has been the subject of considerable interest from both theorists and experimentalists; the fact that a protein typically adopts a single structure is in contrast to most polymers, where individual monomers can adopt a variety of conformers.

In addition to explaining biological function, the study of peptides and proteins has led to the field of molecular self-assembly. Defined as the spontaneous organisation of individual components into an ordered structure without human intervention [12], molecular self-assembly is ever present in nature and has emerged as a new strategy in chemical synthesis, polymer science and engineering [13]. The challenge being to design molecular building blocks that can undergo this spontaneous organisation, using non-covalent bonds such as hydrophobic interactions, hydrogen bonds and electrostatic interactions [14]. While these forces are very week separately, combined they are strong enough to drive the self-assembly process. A major focus of nanotechnology has been the attempt to use biological molecules in constructing supramolecular structures [15, 16], with peptides being exploited in forming ribbons [17], nanotubes [18, 19], and monolayers with nanoscale order [20]. As such, the design of small peptide-based molecules with extended β-strand conformation that self-assemble has been an active area of research [21]. Furthermore, a higher order of self-assembly leads to the formation of fibrils and gels [22] - not only is this important in the design of biomaterials, but it is also useful in studying pathogenesis of certain age-related diseases; many fatal neurodegenerative diseases such as Alzheimer's, Parkinson's, and prion related diseases are associated with the self-assembly of mis-folded proteins (or polymer fragments) into amyloid fibrils [23–25].

In order to probe structural variations in proteins (rationalising biological phenomena such as antibody discrimination, viral fusion or fibril formation), biophysicists seek to study such reactions from the perspective of physical sciences by stripping the complex reactions down to the most basic elements. To aid in the investigation into protein folding, this doctoral study has made use of a combination of computer simulations and experimental techniques to investigate a selection of biologically relevant peptides. This has comprised of utilising classical atomistic molecular dynamics (MD) simulations to characterise the free-energy landscapes of the chosen peptides, and compare these findings with the secondary structure content predicted by spectroscopic methods such as circular dichroism and infrared spectroscopy. As MD can provide information about the conformational properties of molecular systems and how the conformation changes with time, such a combination of techniques should allow one to probe any changes to this structure arising from variations in the environment or mutations introduced at the sequence level. While this work sought to break down the complex biological phenomena into

simple models, a major driving force was to ensure that the models were representative of real biological systems. To that end, a collection of peptides found in hormones, viruses and amyloid diseases were selected in order to probe a variety of questions. Namely, how the primary sequence influenced the three-dimensional structure (and thus its biological function), how such a conformation could be influenced by its environment, and how these systems aggregated.

1.2 Thesis Layout

In order to understand the methods used throughout this work, *Chapter 2* introduces the classical molecular dynamics simulation methodology for the benefit of those with no prior exposure to this field. This introduction covers how the trajectories of atoms are computed, the forces acting on said atoms, and some tricks of the trade. As part of the attempt to compare and contrast results obtained from such simulations to those obtained experimentally, *Chapter 2* also describes the experimental techniques used.

Chapter 3, introduces the transmembrane HIV fusion peptide $gp41_{659-671}$ and probes how the environment can induce variations in the secondary structure of this peptide. This is achieved by predicting the structure of $gp41_{659-671}$ using a variety of classical atomistic force fields and through solvents mimicking the various environmental conditions experienced in its native environment.

In *Chapter 4*, the interactions of a dimeric system $gp41_{532-544}$ and $gp41_{659-671}$ are investigated under both aqueous and mixed lipidic conditions in order to probe structural variations induced during the hemifusion phase of viral infection. Computer simulations are exploited to probe if the secondary structure of each peptide is affected by the presence of the other, while simultaneously ascertaining the tertiary contacts of this system in comparison to experimental data.

Chapter 5 is concerned with the structural determination of the human chorionic gonadotropin (hCG) hormone using both experimental and computer simulation techniques. This chapter also covers an investigation into structural variation between hCG, the near identical luteinizing hormone (LH) and a variety of other individual amino-acid mutations, utilising molecular dynamic simulations. Particular emphasis is placed on the effect a point mutation has on finer

structural elements of the peptides.

In *Chapter 6*, the structure of various oligomers of the amyloid fibril forming human transthyretin (TTR) fragment $TTR_{105-115}$ are studied utilising gas-phase molecular dynamics simulations. The results of which are used to compute collisional cross-sectional areas which are subsequently compared to values obtained previously from ion-mobility mass spectrometry (IM-MS), in order to gain further insight into the early aggregation states.

2

Simulation and Experimental Methodology

In this chapter, the basics of Molecular Dynamics will be introduced for the benefit of individuals with no prior exposure to the field. Further reading in the form of Allen & Tildesley [26], Leach [27] and Frenkel & Smit [28] is recommended. The latter half of the chapter focuses on the experimental techniques employed throughout the thesis.

2.1 Molecular Dynamics

Before discussing the theory behind the computer simulations, it is important to consider why such simulations are conducted. Primarily, they allow one to study properties of a many-particle system by generating information at the microscopic level (atomic positions etc.) and convert this into macroscopic terms (energy, pressure etc.). This is because some of the detailed microscopic information such as instantaneous positions and velocities cannot be directly compared to experimental data, which usually measures a property averaged over numerous particles and time. It is therefore important to consider which averages such simulations can compute, therefore a brief introduction to Statistical Mechanics is necessary. This branch of physics concerns the calculation of thermodynamic properties via ensemble averages, and it can be shown that such averages are themselves directly comparable to averages acquired by measuring a quantity over a specific time period, which is the basis of Molecular Dynamics simulations.

2.1.1 Statistical Mechanics

The 1^{st} and 2^{nd} laws of thermodynamics state that the energy of the universe is conserved and that the entropy of the universe only increases. This results in an expression for an infinitesimal change in energy dE, which relates the energy input due to heating, TdS; where T is the temperature, and dS is the change in entropy; and the work done, PdV; where P is the pressure and dV is the change in volume (Eq. 2.1). This can be expanded for an open system of variable particle number, dN, of chemical potential μ (Eq. 2.2).

$$dE = TdS - PdV \tag{2.1}$$

$$dE = TdS - PdV + \mu dN \tag{2.2}$$

In a theoretical system of energy E with two weakly interaction subsystems of E_1 and E_2 (fixed volume and particle number), it is found that while E_1 and E_2 may fluctuate (i.e. $E_1 \rightarrow E_1 + dE_1$, $E_2 \rightarrow E_2 + dE_2$), the total energy is fixed ($E_2 + dE_2 = E_2 - dE_1$). Furthermore, the change in total entropy ($S = S_1 + S_2$) is found to be zero at equilibrium (Eq. 2.3) as the entropy is at a maximum.

$$\begin{aligned} dS = dS_1 + dS_2 &= \frac{\partial S_1}{\partial E_1} dE_1 + \frac{\partial S_2}{\partial E_2} dE_2 \\ &= \left(\frac{\partial S_1}{\partial E_1} - \frac{\partial S_2}{\partial E_2} \right) dE_1 \\ &= 0 \end{aligned} \tag{2.3}$$

$$\frac{\partial S_1}{\partial E_1} = \frac{\partial S_2}{\partial E_2} \tag{2.4}$$

The 0^{th} law of thermodynamics tells us that if the two subsystems are in equilibrium (Eq. 2.4) they must have the same temperature, T. This results in the thermodynamic definition of temperature (Eq. 2.5) where V and N are fixed.

$$\frac{\partial S}{\partial E} = \frac{1}{T} \tag{2.5}$$

By rearranging the expression for an infinitesimal change in energy dE (Eq. 2.2), in favour of dS (Eq. 2.6), one can see that such a definition can be obtained directly. Furthermore, additional fundamental properties can be be obtained (Eq. 2.7).

$$dS = \frac{1}{T}dE + \frac{p}{T}dV - \frac{\mu}{T}dN \tag{2.6}$$

$$\begin{aligned} \frac{1}{T} &= \left(\frac{\partial S}{\partial E}\right)_{V,N} \\ \frac{p}{T} &= \left(\frac{\partial S}{\partial V}\right)_{E,N} \\ \frac{\mu}{T} &= -\left(\frac{\partial S}{\partial N}\right)_{E,V} \end{aligned} \tag{2.7}$$

2.1.1.1 Quantum Mechanical

The basic laws of Statistical Mechanics can be easily derived using a basic knowledge of quantum mechanics. All that is required is the understanding that a quantum mechanical system can be found in different states, and for the purposes of this discussion such states ($|i>$) are limited to those that are eigenvectors of the Hamiltonian $\mathcal{H}$ of a system. That is to say, for any of these states, $\mathcal{H}|i> = E_i|i>$ where E_i is the energy of that state. Furthermore, the number of eigenstates with energy E in a system of volume V and N particles, may be denoted $\Omega(E, V, N)$. Finally, in a system with fixed E, V, and N, the system is likely to be found in any of its eigenstates $\Omega(E)$.

In the theoretical system of two weakly interaction subsystems, for any given value of E_1, $\Omega(E_1) \times \Omega(E_2)$ gives the total number of degenerate states - however it is more convenient to express this in a form utilising the natural logarithms (Eq: 2.8).

$$\ln \Omega(E_1, E - E_1) = \ln \Omega_1(E_1) + \ln \Omega_2(E - E_1) \tag{2.8}$$

One may then evaluate the most likely value of E_1 as the value that maximises $\ln \Omega(E_1, E - E_1)$ (Eq. 2.9).

$$\left(\frac{\partial \ln \Omega(E_1, E - E_1)}{\partial E_1}\right)_{N,V,E} = 0$$

$$\left(\frac{\partial \ln \Omega_1(E_1)}{\partial E_1}\right)_{N_1,V_1} = \left(\frac{\partial \ln \Omega_2(E_2)}{\partial E_2}\right)_{N_2,V_2} \tag{2.9}$$

By introducing a shorthand notation (Eq. 2.10), this can be redefined as $\beta(E_1, V_1, N_1) = \beta(E_2, V_2, N_2)$. These two subsystems are therefore in (thermal) equilibrium when this condition is met and there is no net energy flow between them.

$$\beta(E, V, N) \equiv \left(\frac{\partial \ln \Omega(E, V, N)}{\partial E}\right)_{N,V} \tag{2.10}$$

The second law of thermodynamics states that the entropy, S, of a system is at a maximum when at thermal equilibrium, and can be linked to $\ln \Omega$ via the Planck relationship using the Boltzmann's constant k_B (Eq. 2.11).

$$S(E, V, N) \equiv k_B \ln \Omega(E, V, N) \tag{2.11}$$

Exploiting the fact that two bodies in thermal equilibrium have the same temperature, T, one can use the thermodynamic definition of temperature (Eq. 2.5) to define a relationship between β and T (Eq. 2.12).

$$\beta = \frac{1}{k_B T} \tag{2.12}$$

If system in a heat bath, b, has a state i with energy E_i, the bath therefore has energy $E_b = E - E_i$ and degeneracy $\Omega_b(E - E_i)$. The probability of finding system 1 in state i is dependent on this degeneracy (Eq. 2.13), and can be evaluated by using a rearranged Planck relationship (Eq. 2.14) and a taylor expansion assuming $E_i \ll E$ (Eq. 2.15).

$$P_i = \frac{\Omega_b(E - E_i)}{\sum_j \Omega_b(E - E_i)} \tag{2.13}$$

$$\Omega_b(E - E_i) = exp\left[\frac{S_b(E - E_i)}{k_B}\right] \tag{2.14}$$

$$S_b(E - E_i) = S_b(E) - E_i \frac{\partial S_b}{\partial E_b} + \frac{E_i^2}{2} \frac{\partial^2 S_b}{\partial E_b^2} \tag{2.15}$$

The 3^{rd} term of this taylor expansion can be shown to tend to zero as it relates to the heat capacity of the bath, $C_b = \partial E/\partial T$ (Eq. 2.16), where it is assumed that the bath can transfer energy to the system without changing the temperature (i.e. $C_b \to \infty$).

$$\frac{E_i^2}{2}\frac{\partial^2 S_b}{\partial E_b^2} = \frac{E_i^2}{2}\frac{\partial}{\partial E_b}\left(\frac{1}{T}\right) = -\frac{E_i^2}{2T^2}\frac{\partial T}{\partial E_b} = -\frac{E_i^2}{2T^2}\left[\frac{1}{C_b}\right] \tag{2.16}$$

Setting $S_b(E)$ as a constant, results in the well-known Boltzmann distribution (Eq. 2.17) which can be used to compute average values such as energy, $\langle E \rangle$, at a given temperature (Eq. 2.18).

$$P_i = \frac{exp(-E_i/k_BT)}{\sum_j exp(-E_j/k_BT)} \tag{2.17}$$

$$\begin{aligned}\langle E \rangle &= \sum_i E_i P_i \\ &= \frac{\sum_i E_i exp(-E_i/k_BT)}{\sum_j exp(-E_j/k_BT)} \\ &= \frac{1}{\mathcal{Z}}\sum_i E_i exp(-E_i/k_BT)\end{aligned} \tag{2.18}$$

This average makes use of the shorthand notation $\mathcal{Z} = \sum_j exp(-E_j/k_BT)$ which is known as the partition function.

2.1.1.2 Classical

If one wished to generalise the relationship for the average energy $\langle E \rangle$ to find the average of some observable $\mathcal{A}$, one would use the expectation value of the operator $\mathcal{A}$ in quantum state i ($< i|\mathcal{A}|i >$) in conjunction with the probability that a system at temperature T being found in an energy eigenstate E_i (Eq. 2.19).

$$\langle \mathcal{A} \rangle = \frac{\sum_i exp(-E_i/k_BT) < i|\mathcal{A}|i >}{\sum_j exp(-E_j/k_BT)} \tag{2.19}$$

Solving such a relationship would appear to be difficult, as it suggests that one solve the Schrödinger equation for a many-body system. Furthermore if this was indeed feasible, the

number of resulting quantum states would be too large to evaluate. As such, an alternative method should be sought. In *Section 2.1.1.1*, it was noted that $\mathcal{H}|i>=E_i|i>$, therefore one can use the Hamiltonian, $\mathcal{H}$, as the basis for an expression relating to $-E_i/k_BT$ (Eq. 2.20) which in turn could be incorporated into the expression for $\langle\mathcal{A}\rangle$ (Eq. 2.21).

$$exp(-E_i/k_BT) =< i|exp(-\mathcal{H}/k_BT)|i > \tag{2.20}$$

$$\begin{aligned}\langle\mathcal{A}\rangle &= \frac{\sum_i < i|exp(-\mathcal{H}/k_BT)\mathcal{A}|i >}{\sum_j < j|exp(-\mathcal{H}/k_BT)|j >} \\ &= \frac{Tr\ exp(-\mathcal{H}/k_BT)\mathcal{A}}{Tr\ exp(-\mathcal{H}/k_BT)}\end{aligned} \tag{2.21}$$

The trace of a matrix, Tr, is the sum of the (complex) eigenvalues and is invariant with respect to a change of basis set; if Λ is a n-by-n matrix with real or complex entries, and $\lambda_1 \ldots \lambda_n$ are the (complex) eigenvalues of Λ, then the trace of the matrix is $Tr(\Lambda) = \sum \lambda_i$.

The Hamiltonian operator, $\mathcal{H}$, is itself a sum of the kinetic energy operator, $\mathcal{K}$, and the potential energy operator, $\mathcal{U}$. The former relates to the momenta of particles ($\mathcal{K} = p^2/m$) and as such, momentum eigenstates are eigenfunctions of this operator. The later relies on the positions of all particles, therefore the ideal basis set would be constructed around position eigenfunctions. Unfortunately, in order to exploit the use of eigenvectors and eigenvalues, $\mathcal{H} = \mathcal{K} + \mathcal{U}$ and $exp[-\beta(\mathcal{K}+\mathcal{U})]$ would need to be diagonal in either basis set, which they are not. Furthermore, one can not simply replace $exp(-\beta\mathcal{H})$ with $exp(-\beta\mathcal{K})exp(-\beta\mathcal{U})$ due to the commutator $[\mathcal{K},\mathcal{U}]$ (Eq. 2.22).

$$exp(-\beta\mathcal{K})exp(-\beta\mathcal{U}) = exp\{-\beta[\mathcal{K}+\mathcal{U}+\mathcal{O}([\mathcal{K},\mathcal{U}])]\} \tag{2.22}$$

A commutator $[a,b]$ indicates if a binary operation is not commutative, i.e. changing the order changes the end result; $[a,b] = ab - ba \neq 0$. Applying this in the field of quantum mechanics, $[\mathcal{K},\mathcal{U}] = -i\hbar$, where $\hbar = h/2\pi$ and h is Planck's constant (Eq. 2.23).

$$\begin{aligned}[\mathcal{K},\mathcal{U}]\Psi &= (\mathcal{K}\mathcal{U} - \mathcal{U}\mathcal{K})\Psi \\ &= -i\hbar\frac{\partial}{\partial x}(x\Psi) + ix\hbar\frac{\partial}{\partial x}(\Psi) \\ &= -i\hbar\Psi - ix\hbar\frac{\partial\Psi}{\partial x} + ix\hbar\frac{\partial\Psi}{\partial x} \\ &= (-i\hbar)\Psi \end{aligned} \tag{2.23}$$

However in the limit $\hbar \to 0$, this term can be ignored (Eq. 2.24) and $exp(-\beta\mathcal{H})$ can be approximated in terms of exponentials of $\mathcal{K}$ and $\mathcal{U}$ (Eq. 2.25).

$$\begin{aligned} exp(-\beta\mathcal{K})exp(-\beta\mathcal{U}) &= exp\{-\beta[\mathcal{K}+\mathcal{U}+\mathcal{O}([\mathcal{K},\mathcal{U}])]\} \\ &\approx exp\{-\beta[\mathcal{K}+\mathcal{U}]\} \end{aligned} \tag{2.24}$$

$$\begin{aligned} Tr\ exp(-\beta\mathcal{H}) &= Tr\ exp[-\beta(\mathcal{K}+\mathcal{U})] \\ &\approx Tr\ exp(-\beta\mathcal{U})exp(-\beta\mathcal{K}) \end{aligned} \tag{2.25}$$

This approach results in directly evaluable terms for the potential energy operator, $\mathcal{U}$ (Eq. 2.26) and the kinetic energy operator, $\mathcal{K}$ (Eq. 2.27). Which in turn can be integrated over all coordinates and momenta of the system (Eq. 2.28), factoring in the indistinguishability of identical particles (1/N!) and that every quantum state corresponds to a volume of h^{dN}, where d is the dimensionality of the system (i.e. h^{3N} for a 3-dimensional system).

$$exp[-\beta\mathcal{U}(\mathbf{r}^N)] \tag{2.26}$$

$$exp\left[-\beta\sum_{i=1}^{N} p_i^2/(2m_i)\right] \tag{2.27}$$

$$Tr\ exp(-\beta\mathcal{H}) \approx \frac{1}{h^{dN}N!}\int d\mathbf{p}^N d\mathbf{r}^N\ exp\left\{-\beta\left[\sum_i p_i^2/(2m_i) + \mathcal{U}(\mathbf{r}^N)\right]\right\}$$
$$\equiv \mathcal{Z}_{classical} \tag{2.28}$$

Such an approach can also be applied to derive an expression for $Tr\ exp(-\beta\mathcal{H})\mathcal{A}$, which allows one to evaluate the classical average of $\langle\mathcal{A}\rangle$ (Eq. 2.29).

$$\langle\mathcal{A}\rangle = \frac{\int d\mathbf{p}^N d\mathbf{r}^N exp\{-\beta[\sum_i p_i^2/(2m_i) + \mathcal{U}(\mathbf{r}^N)]\}\mathcal{A}(\mathbf{p}^N, \mathbf{q}^N)}{\int d\mathbf{p}^N d\mathbf{r}^N exp\{-\beta[\sum_j p_j^2/(2m_j) + \mathcal{U}(\mathbf{r}^N)]\}} \tag{2.29}$$

2.1.2 Ergodicity

So far, the expected value of an observable $\mathcal{A}$ has been shown to be obtained from an ensemble average using classical statistical mechanics. However within the context of experimental measurement, one does not typically measure an average over all possible states. Instead, one typically measures a quantity over a specific period of time, then averages accordingly. Properties such as pressure and density depend upon the positions (**r**) and momenta (**p**) of the many particles with a system, which will naturally vary over time due to interactions between particles. The experimentally acquired average is therefore the average of the instantaneous value $\mathcal{A}(\mathbf{p}^N(t), \mathbf{r}^N(t))$ over the time of the measurement. Intuitively, as this period of time increases to infinity (Eq. 2.30), one would expect the average measurement to approach an acceptable 'true' reflection of the property.

$$\overline{\mathcal{A}} = \lim_{t\to\infty}\frac{1}{t}\int_0^t dt'\mathcal{A}(\mathbf{p}^N(t'), \mathbf{r}^N(t')) \tag{2.30}$$

The density of a fluid (ρ) is a prime example of a time dependent quantity of any system. At a distance r from an atom i, the density $\rho_i(r)$ will depend on the coordinates of all particles in the system which will naturally vary over time. Once can therefore evaluate the time evolution of the density if the initial coordinates and momenta of all particles are known; $\rho_i(r; \mathbf{r}^N(0), \mathbf{p}^N(0), t)$. Furthermore, one could assume that over a significantly large time period, the average $\overline{\rho_i(r)}$ would become independent of the initial coordinates and momenta (Eq. 2.31).

$$\overline{\rho_i(r)} = \lim_{t\to\infty} \frac{1}{t} \int_0^t dt' \rho_i(r;t') \tag{2.31}$$

While this may not strictly be true for finite time periods, in the case of a macroscopic experiment it can be assumed that the result of $\overline{\rho_i(r)}$ would be independent of the initial coordinates (I.C.). As such, one would expect the result of $\overline{\rho_i(r)}$ to be same as if we averaged over a large number of systems of disparate I.C. ($N_{I.C.}$) and momenta (Eq. 2.32), but with identical values of N, V, and E (Eq. 2.33).

$$\overline{\rho_i(r)} = \frac{1}{N_{I.C.}} \sum_{I.C.} (\lim_{t\to\infty} \frac{1}{t} \int_0^t dt' \rho_i(r; \mathbf{r}^N(0), \mathbf{p}^N(0), t')) \tag{2.32}$$

$$\frac{1}{N_{I.C.}} \sum_{I.C.} f(\mathbf{r}^N(0), \mathbf{p}^N(0)) \to \frac{\int_E d\mathbf{r}^N d\mathbf{p}^N f(\mathbf{r}^N(0), \mathbf{p}^N(0))}{\Omega(N,V,E)} \tag{2.33}$$

From the previous section, such an ensemble average is typically expressed as $\langle f \rangle$, which results in a simplification in the expression for $\overline{\rho_i(r)}$ (Eq. 2.34).

$$\overline{\rho_i(r)} = \lim_{t\to\infty} \frac{1}{t} \int dt' \langle \rho_i(r; \mathbf{r}^N(0), \mathbf{p}^N(0), t') \rangle_{NVE} \tag{2.34}$$

The ensemble average can be regarded as being time independent and averaging over the initial phase space coordinates would naturally be the same as averaging over the phase space coordinates at a later time. This results in a simple relationship between the time averaged $\overline{\rho_i(r)}$ and the ensemble averaged $\langle \rho_i(r) \rangle_{NVE}$ (Eq. 2.35).

$$\overline{\rho_i(r)} = \langle \rho_i(r) \rangle_{NVE} \tag{2.35}$$

If one therefore wishes to calculate the average of a function dependent on the coordinates and momenta of a many-body system, then one can either do so by means of an 'ensemble' or 'time' average. The former is typically the approach of Monte Carlo Simulations, while the later is the basis of Molecular Dynamics which is used extensively throughout this body of work.

2.1.3 General Classical Molecular Dynamics

Molecular Dynamics (MD) simulations essentially investigate motions of a system of discrete particles which are influenced by internal and external forces. The underlying principles of MD are that the interactions of said particles are empirically described by a potential energy function, from which may be derived the forces acting on each particle. Successive configurations of the system of particles are subsequently generated by integrating Newton's laws of motion. These laws state that in the absence of an external field, a body will continue to move in a straight line and for every force there is a corresponding equal and opposite force. Furthermore, they explicitly equate the force acting on an object to the subsequent change in momentum that the object will experience (Eq. 2.36).

$$m_i \frac{\partial^2 \mathbf{r}_i}{\partial t^2} = \mathbf{f}_i, i = 1...n \tag{2.36}$$

One can calculate the trajectories of the atoms by solving the differential equations associated with this relationship.

2.1.3.1 Newtonian Dynamics & Basic Algorithms

As the positions of particles varies with time, the interactions between particles will also change. The force is therefore dependent on the particles position relative to its neighbours, and as this position fluctuates so will the force. It is therefore important to break down the integration into small time steps of δt. The procedure is therefore to evaluate the force on each particle at time t on the basis of its interactions with other particles. The acceleration is the derived from the force and, in combination with the position at time t, is used to calculate the positions of the particles at time $t + \delta t$. It should be noted that the assumption made is that the force is invariant over the course of δt From these new positions, the new force on each particle is evaluated on order to generate a new set of positions at time $t + 2\delta t$ etc.

Several algorithms have been developed to integrate the Newtonian equations of motion, which assume that the dynamic properties can be treated as a Taylor expansion about δt (Eq. 2.37).

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \dot{\mathbf{r}}(t)\delta t + \frac{\ddot{\mathbf{r}}(t)}{2}(\delta t)^2 + \frac{\dddot{\mathbf{r}}(t)}{6}(\delta t)^3 + ...$$
$$\mathbf{r}(t-\delta t) = \mathbf{r}(t) - \dot{\mathbf{r}}(t)\delta t + \frac{\ddot{\mathbf{r}}(t)}{2}(\delta t)^2 - \frac{\dddot{\mathbf{r}}(t)}{6}(\delta t)^3 + ... \tag{2.37}$$

Combination of the Taylor expansions for $\mathbf{r}(t+\delta t)$ and $\mathbf{r}(t-\delta t)$ results in an expression for $\mathbf{r}(t+\delta t)$ based on the positions and accelerations at time t (i.e. $\mathbf{r}(t)$ & $\ddot{\mathbf{r}}(t)$) in conjunction with the positions from the previous step, $\mathbf{r}(t-\delta t)$ (Eq. 2.38). This is the basis of the Verlet algorithm and is perhaps the most widely used method for integrating the equations of motion.

$$\mathbf{r}(t+\delta t) + \mathbf{r}(t-\delta t) = 2\mathbf{r}(t) + \ddot{\mathbf{r}}(t)(\delta t)^2 + \mathcal{O}(\delta t)^4$$
$$\mathbf{r}(t+\delta t) \approx 2\mathbf{r}(t) - \mathbf{r}(t-\delta t) + \ddot{\mathbf{r}}(t)(\delta t)^2 \tag{2.38}$$

An important feature of the Verlet algorithm relates to the symmetrical roles of $r(t+\delta t)$ and $r(t-\delta t)$, resulting in it being time-reversible (which is an accurate representation of Newton's equations of motion). Another notable feature is that the velocities $\dot{\mathbf{r}}$ are not required in order to compute the trajectories of the particles, as they have been eliminated as a consequence of the summation of the Taylor expansions about **r**(t). If so desired, one can evaluate the velocities $\dot{\mathbf{r}}(t)$ in a variety of ways with a simple approach being a manipulation of the Taylor expansion for $\mathbf{r}(t+\delta t)$ and $\mathbf{r}(t-\delta t)$ (Eq. 2.39).

$$\mathbf{r}(t+\delta t) - \mathbf{r}(t-\delta t) = 2\dot{\mathbf{r}}(t)\delta t + \mathcal{O}(\delta t)^3$$
$$\dot{\mathbf{r}}(t) = \frac{\mathbf{r}(t+\delta t) - \mathbf{r}(t-\delta t) + O(\delta t)^3}{2\delta t} \approx \frac{\mathbf{r}(t+\delta t) - \mathbf{r}(t-\delta t)}{2\delta t} \tag{2.39}$$

Even though the velocities are not necessary in order to advance the positions of the particles, they are useful for estimating the Kinetic energy ($\mathcal{K}$) and hence the total energy of the system ($\mathcal{H} = \mathcal{K} + \mathcal{U}$). A major drawback evident from this is that the velocities are only available after the positions have been calculated for the next time step. Other potential drawbacks of the Verlet algorithm lie in the approximation of $\mathbf{r}(t+\delta t)$ and $\dot{\mathbf{r}}(t)$. In these expressions there are associated errors of $(\delta t)^4$ and $(\delta t)^2$ respectively. Additionally at time $t = 0$ there is no information regarding $\mathbf{r}(-\delta t)$, necessary for the generation of $\mathbf{r}(\delta t)$. This can be circumnavigated by truncating

the Taylor expansion; $\mathbf{r}(-\delta t) = \mathbf{r}(0) - \delta t\dot{\mathbf{r}}(0)$. Several other variations of the Verlet algorithm have been developed in order to tackle some of the deficiencies associated with the algorithm. One such example is the leap frog algorithm which evaluates the positions at time $\mathbf{r}(t + \delta t)$ and velocities at $\dot{\mathbf{r}}(t + \frac{1}{2}\delta t)$. The major weakness of this algorithm is its inability to compute these quantities at the same time, resulting in the kinetic and potential energies being undefined at the same time.

The final algorithm of note is the so-called Velocity Verlet algorithm. This algorithm makes use of both positions and velocities calculated at equal times, with the new velocities being evaluated after the new positions and forces have been calculated (Eq. 2.40).

$$\begin{aligned}\mathbf{r}(t + \delta t) &= \mathbf{r}(t) + \dot{\mathbf{r}}(t)\delta t + \frac{\ddot{\mathbf{r}}(t)}{2}(\delta t)^2 \\ \dot{\mathbf{r}}(t + \delta t) &= \dot{\mathbf{r}}(t) + \frac{\ddot{\mathbf{r}}(t + \delta t) + \ddot{\mathbf{r}}(t)}{2}\delta t\end{aligned} \tag{2.40}$$

2.1.4 Force Field

As mentioned above, the essential features of a MD simulation revolve around the step by step evaluation of the classical equations of motion. The forces acting on atoms, $\mathbf{f}_i$, are usually derived from a potential energy $\mathcal{U}(\mathbf{r}^N)$ where $\mathbf{r}^N = (\mathbf{r}_1, \mathbf{r}_2 ... \mathbf{r}_N)$ represents the complete set of 3N atomic coordinates. The potential energy itself can be regarded as being made up from contributions attributed to two types of interaction, these being non-bonded & bonded.

2.1.4.1 Non-Bonded Interaction

The part of the potential energy which represents contributions from non-bonded interactions, $\mathcal{U}_{n.b.}$, is traditionally split into 1-body, 2-body, 3-body etc. terms (Eq. 2.41).

$$\mathcal{U}_{n.b.}(\mathbf{r}^N) = \sum_i \nu_1(\mathbf{r}_i) + \sum_i \sum_{j>i} \nu_2(\mathbf{r}_i, \mathbf{r}_j) + \sum_i \sum_{j>i} \sum_{k>j>i} \nu_3(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) + ... \tag{2.41}$$

where $\sum_i \sum_{j>i}$ represents the summation over distinct pairs, ensuring no double counting occurs (the same care is taken for triplets etc). The initial 1-body term may represent either an externally applied field or the effects of the container wall. In the case of periodic simulations of

bulk materials, this term is redundant and therefore dropped. The remaining terms, ν_2, ν_3, etc., represent particle interaction. 4-body terms (& higher) are expected to be small in comparison to ν_2 & ν_3 and are therefore neglected. It is also common practise to neglect the 3-body term and focus on the 2-body interactions. This is due in part to the computation time involved in evaluating the 3-body effects. In order to compensate for this, the effects from 3-body interactions may be included in an 'effective pair potential' ν_2^{eff}. The noticeable difference in this 'effective' potential is that it may be dependent on the density or temperature of the system involved, whereas a true 2-body potential is not. At this point it should be noted that as the pair potential only depends on the magnitude of the pair separation $r_{ij} = |\mathbf{r}_j - \mathbf{r}_i|$, one may write the potential as $\nu_2(r_{ij})$. The most commonly used potential is that of the Lennard-Jones interaction (Eq. 2.42) which is defined in terms of the diameter, σ, and the well depth, ϵ_{ij}.

$$\nu^{LJ}(r) = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \tag{2.42}$$

The exact value of these two parameters is material dependent and in the instances where the material in question is not monoatomic, the are approximated according to the Lorentz-Berthelot relationships (Eq. 2.43 & 2.44).

$$\sigma_{ij} = \frac{1}{2}(\sigma_{ii} + \sigma_{jj}) \tag{2.43}$$

$$\epsilon_{ij} = \sqrt{\epsilon_{ii}\epsilon_{jj}} \tag{2.44}$$

Finally, if electrostatic charges are present then interactions are modelled through the coulombic potential (Eq. 2.45), which is defined in terms of the charges, q_1 & q_2, the relative dielectric constant, ϵ, and the permittivity of free space, ϵ_o; ϵ is set to one in calculations with explicit solvents [29]. This is a long range interaction, and can be successfully handled via the Ewald summation method (*Appendix B*) [26–28].

$$\nu^{coulomb}(r) = \frac{q_1 q_2}{4\pi\epsilon_o\epsilon r_{12}} \tag{2.45}$$

2.1.4.2 Bonded Interaction

So far we have only concerned ourselves with the interactions of monoatomic systems. However in the case of molecules we must also consider intramolecular bonding interactions (Eq. 2.46). One can follow these through a molecules equilibrium bond-lengths and angles etc. By treating bonds (and also angles) as pseudo-springs, one can assign a potential energy to deviations from equilibrium (*cf* Hookes Law).

$$\begin{aligned}\mathcal{U}_{intra} = &\frac{1}{2}\sum_{bonds} k^r_{ij}(r_{ij} - r_{eq})^2 \\ &+ \frac{1}{2}\sum_{angles} k^\theta_{ijk}(\theta_{ijk} - \theta_{eq})^2 \\ &+ \frac{1}{2}\sum_{proper}\sum_{m} k^{\phi,m}_{ijkl}\big(1 + \cos(m\phi_{ijkl} - \phi_0)\big) \\ &+ \frac{1}{2}\sum_{improper} k^\xi_{ijkl}(\xi_{ijkl} - \xi_{eq})^2 \end{aligned} \tag{2.46}$$

The first term, $\sum_{bonds}$, relates to the bond stretching between two covalently bonded atoms i and j. The second term, $\sum_{angles}$, is concerned with the bond-angle vibration between the triplet of atoms i-j-k (where j is the centre atom) and is defined by the angle θ_{ijk} (Eq. 2.47).

$$\theta_{ijk} = \cos^{-1}\frac{\mathbf{r}_{ij}.\mathbf{r}_{kj}}{r_{ij}r_{kj}} \tag{2.47}$$

This triplet of atoms may be further restrained by use of an additional potential function known as the Urey-Bradley potential (Eq. 2.48). Taking a similar form to $\sum_{bonds}$, it acts by adding a virtual (or 'ghost') bond between atoms i and k.

$$\mathcal{U}_{UB} = \frac{1}{2}\sum k^{UB}_{ik}\big(r_{ik} - r^{UB}_{eq}\big)^2 \tag{2.48}$$

The last two terms, $\sum_{proper}$ and $\sum_{improper}$, relate to the torsional potentials and are defined by a quartet of atoms i-j-k-l (Fig. 2.1). The proper, or 'dihedral angle potential', is used to restrict rotation around a bond, with the torsional angle ϕ_{ijkl} defined as the angle between the i-j-k and j-k-l planes (Fig. 2.1a). The angle equating to the potentials minimum value is denoted ϕ_0, however the potential may posses numerous minima and the number of which (multiplicity)

is denoted m. The improper term is used to restrict groups of atoms to within a plane as in the case of aromatic rings (Fig. 2.1b) or to retain chirality (Fig. 2.1c), and the angle between the i-j-k and j-k-l planes is denoted ξ_{ijkl}. Such a potential generally only has one minima, therefore is expressed as a harmonic potential akin to $\sum_{bonds}$ and $\sum_{angles}$.

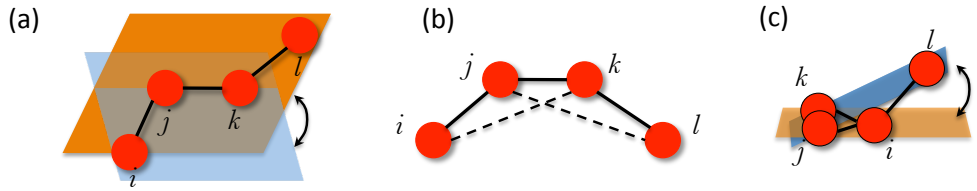


Figure 2.1: Illustration of (a) torsional angle ϕ_{ijkl} in proper dihedral angle potential, (b) restriction of atoms to within a plane in the improper potential, and (c) angle ξ_{ijkl} in improper torsion.

The force acting on an individual atom can therefore be deduced by evaluating all possible potential energy terms associated with it. In order to achieve this, one must know which parameters to use for the numerous atomic combinations and environments in each of the terms. This information is usually stored in a collective database known as a 'force field.' It should be noted that while CHARMM force fields (introduced in *Section 2.1.4.3*) generally makes use of improper dihedrals, Urey-Bradley potentials are only used in special cases [29].

2.1.4.3 Which Force Field?

An important feature to consider in setting up a molecular dynamics simulation is the choice of force field itself. This choice not only defines the functional form of the potential energy, but also defines the various parameters (e.g. σ, q, r_{eq}) used in its evaluation. It is important to note that one may encounter two such force fields with identical functional forms yet each comprised of a unique set of values for these parameters. Interestingly, regardless of completely disparate parameters, these two force fields may be of comparable accuracy. Furthermore, an additional force field defined by a different functional form may also yield comparable accuracy.

Within the force field itself, it is not uncommon to find various types of the same element listed with different parameters. The rationale for this is that (for example) the carbon atom may be found in nature in a variety of forms (i.e. sp^3, sp^2, sp^2 aromatic, and sp). The equilibrium angle θ_{eq} is expected to be unique in each of these hybridisations; sp^3-hybridised C is tetrahedral

and the angle is 109.5°, whereas sp^2-hybridised C is trigonal therefore resulting in an angle of 120°. The neighbouring atoms are also an important feature to consider, for example the oxygen atom can be bound to hydrogen (e.g. water, alcohol) or carbon (e.g. ketone, ether) and parameters such as bond length and strength are expected to be different in each case. Such parameters are usually based on experimental values, for example the equilibrium bond lengths between two atoms may be obtained from electron diffraction data, whereas the force constant for such a bond may be obtained from IR.

In molecular modelling, force fields are primarily designed to predict the structure properties of molecules. Out of the countless variety of force fields available, AMBER [30] and CHARMM [31] have been designed to study macromolecules such as proteins and polymers and have highly similar functional forms, allowing for the possible use of one force field within the others primary software package; other force fields of note are GROMOS and OPLS. However, there are subtle differences between the two force fields, with CHARMM having additional bonded terms in the form of the four-body quadratic improper potential and the two-body Urey-Bradley term (both absent in AMBER); as mentioned in *Section 2.1.4.2*, the latter is only used in special cases [29]. There are also differences in the treatment of the electrostatic and van der Wall scaling factors of 1-4 non-bonded interactions (i.e. atoms separated by two covalent bonds). CHARMM uses values of 1.0 for both, while AMBER uses 1.2 and 2.0 respectively; a value of 1.0 denotes no modification, while a larger value strengthens these non-bonded interactions.

Several studies have noted that the original AMBER force fields (referred to as ff94 and ff99) did not provide an acceptable energy balance between helical and extended regions of peptide and protein backbones. As such, attempts have been made in order to improve the conformation sampling and two notable examples are ff99SB and ff03. AMBER-ff99SB presents a re-parameterisation of the backbone torsion terms in ff99 and achieves a better balance of the four basic secondary structure elements (i.e. β, α_L, α_R and pPII). The new dihedral parameters were obtained by fitting the energies of multiple conformations of glycine and alanine tetrapeptides to high-level *ab initio* quantum mechanical calculations. The second modification of note is AMBER-ff03, which is also a modified version of ff99. In addition to a modification of the ϕ, ψ backbone torsions in order to decrease the preference for helical conformations, the charges were derived from quantum calculations that used a continuum dielectric to mimic

solvent polarisation.

Over stablisaton of helical content at the expense of other secondary structure motifs is not restricted to AMBER force fields [32]. Observed differences between experimental and classically calculated two-dimensional ϕ/ψ free energy surfaces using the CHARMM22 force field resulted in the CMAP correction term [33, 34]. CMAP is a function of two sequential backbone dihedrals and is a numerical energy correction that transforms the classical energy map to match that of a quantum mechanical calculated map. It should be noted that this study does not make use of CMAP corrections, instead it utilises unmodified CHARM22 [31], AMBER-ff99SB [35], and AMBER-ff03 force fields [36, 37].

In addition one may choose from a variety of water models in order to simulate the molecule in aqueous solution. For example, the explicit model TIP3P (which CHARMM is optimised against) describes the water molecule as a rigid three-charge site molecule of fixed bond length, r_{OH} = 0.9572Å, and angle, $\angle HOH$ = 104.52°, while the TIP4P model (4-site) shifts the charge on the oxygen to a dummy site along the bisector of the HOH angle: resulting in improved electrostatic distribution. Implicit solvents are an alternative to explicit water models, which represent the solvent as an effective mean-field potential and are capable of accurately reproducing experimentally measurable quantities [29, 38]. While such an approach requires significantly less computational time than utilising explicit models, investigations utilising implicit solvents have shown that they appear to overestimate secondary structure and intramolecular hydrogen bonding [38]. Such water models may therefore be inappropriate for studying the structure of the models investigated herein, therefore this body of work has utilised the explicit TIP3P model throughout.

2.1.5 Periodic Boundary Conditions

In order to simulate the bulk phase of a system, yet using only a relatively small number of particles, one may employ periodic boundary conditions to a simulation. The cubic box of particles is replicated throughout space to form a periodic array: for example, a two dimensional cell will be surrounded by 8 identical neighbours (Fig. 2.2), whereas a three-dimensional cell will be subsequently surrounded by 26 identical neighbours. During the course of the simulation, movements of particles in the central original box are replicated across this periodic array; in

reality, the positions of the particles contained within these mirror images are obtained through manipulation of the co-ordinates of the central cell by adding (or subtracting) integral multiples of the length of the box, L (i.e. $x+nL, y+nL, z+nL$). A consequence of this is that if a particle crosses a boundary and leaves its box, a particle from a neighbouring 'mirror image' will enter via the opposite boundary. Therefore one only needs to store the co-ordinates of the central box, and only consider the co-ordinates of the mirror images when a particle crosses such a boundary (Fig. 2.2).

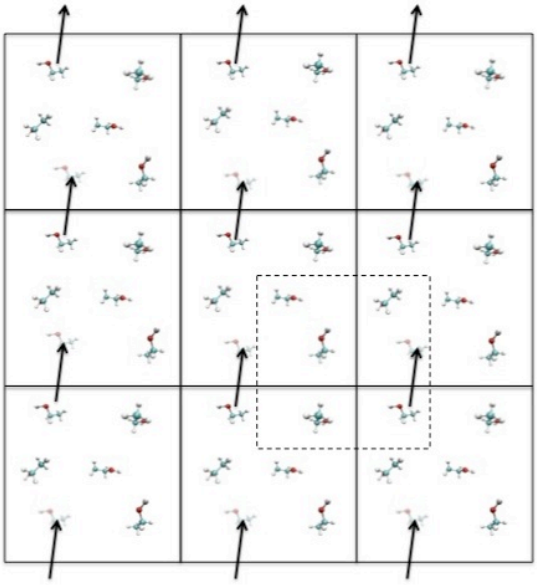


Figure 2.2: Periodic boundary conditions can be visualised as surrounding the central simulation box with an (infinite) number of copies. As atoms reach one edge of the box they immediately appear at the opposite side. The dashed box indicates the limit which a particle can interact with its nearest periodic neighbours (which need not necessarily belong to the same periodic image).

So far only cells based on a cubic geometry have been mentioned. This does not mean that other cell shapes are impossible, in fact the only caveat is that the shape must fill all of space when replicated by translational operations. This is satisfied by five shapes: the aforementioned cube, a hexagonal prism, rhombic dodecahedron, truncated octahedron, and 'elongated' dodecahedron. A parallelepiped cell (closely related to the cubic cell) also fulfils this requirement.

Periodic boundary conditions therefore appear to be a perfect solution to simulating bulk phase properties, however a major problem lies in the question whether the properties of a macro-

scopic system can be accurately represented by a small infinitely periodic system. A long-ranged potential will have considerable contribution from interactions between particles and their own periodic images. Such an interaction is undesirable as it imposes an anisotropic nature to the intended isotropic system. As a consequence of this, several techniques have been developed in order to calculate the potential energy of a system subjected to these periodic boundary conditions. For short-ranged interactions one may exploit the 'mirror image convention' which restricts the summation of the contributions to the potential energy to a region whose size and shape is identical to that of the original central cell (Fig. 2.2). Furthermore, as the largest contribution stems from interactions with a particles 'nearest neighbours,' one may further restrict the summation by applying a spherical cut-off to the potential. This equates to the potential being zero for distances $r \gg r_c$ where r_c is the cut-off; for a cubic cell of dimension L the value of r_c cannot be $> \frac{L}{2}$.

If such a potential is not precisely zero for distances in excess of r_c, then one would expect the total potential energy $\mathcal{U}_{tot}$ to posses systematic errors. This lost energy can be captured by evaluating a tail contribution $\mathcal{U}_{tail}$ in addition to the truncated potential $\mathcal{U}_c$ (Eq. 2.49). This relies on the average number density ρ and assumes that the radial distribution function $g(r)$; the probability density of finding an atom at distance r from another atom at $r = 0$; is unity for distances in excess of the imposed cut-off.

$$\mathcal{U}_{tot} = \mathcal{U}_c + \frac{N\rho}{2} \int_{r_c}^{\infty} dr\, U(r)\, 4\pi r^2 \tag{2.49}$$

Such an approach assumes that the the intermolecular interactions decay rapidly. Clearly, such a statement is true only for potential energy functions which decay faster than r^{-3}, as the $\mathcal{U}_{tail}$ term would be infinite for any potential which decayed slower than this. The Coulombic interaction is a long-ranged potential and decays as r^{-1}, therefore one must find an alternative method that will effectively handle these intermolecular interactions. As mentioned in *Section 2.1.4.1*, this may be achieved via Ewald summation method (*Appendix B*) [26–28].

2.1.6 Parallel Tempering

A major difficulty in simulating complex systems is that the energy landscape of the system is generally very rugged. As a consequence of this, low temperature sampling of the configurational space is often hindered by the inability to cross high-energy barriers. Whilst such simulations will have precise sampling of that particular region of phase space, they are trapped within low energy minima. In order to ensure that a simulation effectively explores the entire configurational space, several techniques such as Parallel Tempering (or Replica Exchange Method) have been developed.

Parallel tempering simulations amount to simulating N replicas of the system of interest in parallel. These are typically simulated in the canonical ensemble with each replica (or 'temperer') at a different temperature ranging from T_1 to T_N, with T_N chosen to ensure that the replica can easily traverse the entire phase space. If so desired, one may replace these different temperatures with another variable thermodynamic property such as pressure or chemical potential. Good sampling of the phase space is achieved by allowing systems at different temperatures to swap their complete configurations at regular intervals during the course of the simulation.

At regular intervals (i.e. after a certain number of time-steps) an attempt is made to swap the configurations of a pair of replicas i and j. This swap between i and j is accepted (or conversely rejected) according to a criterion guaranteeing detailed balance (Eq. 2.50), which is dependent on the energy difference between the two systems, ΔE_{ij}, and the difference in inverse temperatures, $\Delta\beta_{ij}$. From this criteria, it is apparent that swaps between systems become exponentially less likely as the temperature separation increases - thus such swaps usually only occur between adjacent temperers.

$$P_{ij} = min\left\{1, \frac{exp[\Delta\beta_{ij}\Delta E_{ij}]}{1 + exp[\Delta\beta_{ij}\Delta E_{ij}]}\right\} \tag{2.50}$$

The major benefit of running a Parallel Tempering simulations is that configurations sampled by high temperers will eventually swap with those sampled by low temperers. A consequence of this is that unlike in a conventional MD simulation, low temperature replica may jump between local minima separated by high energy barriers. While this body of work does refer to results obtained through parallel tempering, it does not make use of the technique explicitly. Therefore

further information on the topic can be found in literature.

2.2 Experimental Techniques

2.2.1 Circular Dichroism

Circular Dichroism (CD) measures differences in the absorption of left- & right-handed polarised light, arising due to structural asymmetry. The difference in absorption (δA = A_L - A_R) or difference in molar extinction coefficient ($\delta\epsilon$ = ϵ_L - ϵ_R) is recorded at each wavelength, λ, of the absorption band. Since ϵ_L & ϵ_R (or A_L & A_R) typically have nearly the same magnitude, $\delta\epsilon$ may take on either small positive or negative values for optically active compounds, or may be zero as in the case of achiral compounds and racemic mixtures (i.e. equal amounts of left- & right-handed enantiomers).

2.2.1.1 Circular Birefringence & Circular Dichroism

When circularly polarised light passes through an optically active medium, two related phenomena occur. Circular Birefringence occurs when one of the circularly polarised components (Fig. 2.3a) experiences a different refractive index than the other component. This difference in refractive index, δn, is typically 1 millionth of the absolute value of n_L or n_R and results in one of the components being retarded and the linear polarisation vector being rotated by an angle α (Fig. 2.3b). Circular Dichroism occurs when the medium absorbs the two circularly polarised components to a different extent. This results in the superposition of the two components no longer corresponding to a linearly polarised wave but instead tracing out an ellipsoid path (Fig. 2.3c). This superposition is referred to as elliptically polarised light and the degree to which the plane-polarised light becomes elliptical is entirely dependent upon the difference between the absorption's of the two components. In reality it is rare to find a material only exhibiting one of these phenomena, and the resultant superposition comes from a combination of these two effects (Fig. 2.3d).

There exists two requirements for a molecule to exhibit a CD spectrum. Firstly is the presence of a chromophore (a molecule which can absorb radiation through the electronic configuration of its ground state at room temperature). Due to the absorption of energy, the excited state

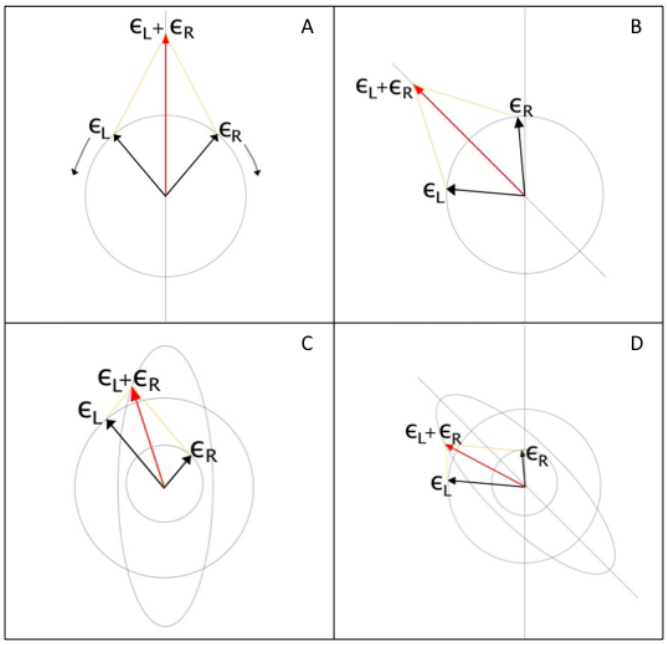


Figure 2.3: Diagrammatic representation of the combination of left, ϵ_L, and right, ϵ_R, circularly polarised light. Resulting in (a) plane-polarised light, (b) rotation of linear polarisation vector as a result of Circular Birefringence, (c) elliptically polarised light as a result of Circular Dichroism, and (d) combination of both phenomena.

interacts with its environment in a unique way which differs from that of its ground state interaction. In the far UV spectral region (190-250nm), the major chromophore is that of a peptide bond. The second requirement for CD is that the chromophore must be in (or closely associated with) an optically asymmetric environment. When a peptide bond is part of a regular folded structure, such as an α-helix or β-sheet, it satisfies this requirement as it is associated with a conformationally asymmetric structure.

2.2.1.2 Mathematical Theory

The absorption of left-handed polarised light can be defined in terms of I_L^0 and I_L, which are the intensities of the light incident on the sample and after travelling a distance l through the sample of molar extinction coefficient ϵ_L and molar concentration C (Eq. 2.51).

$$A_L = log_{10}(I_L^o/I_L) = \epsilon_L C l \tag{2.51}$$

In combination with a similar term for the absorption of right-handed polarised light, one

can derive a relationship between the difference in absorption and difference in molar extinction coefficient (Eq. 2.52).

$$\delta A = A_L - A_R = \epsilon_L Cl - \epsilon_R Cl = (\epsilon_L - \epsilon_R)Cl = \delta\epsilon Cl \tag{2.52}$$

The Beer-Lambert Law (Eq. 2.51) can also be expressed in terms of a natural logarithm (Eq. 2.53), which includes the cross section of light absorption by a single particle σ and the density of absorbing particles N. It should be noted that the absorption coefficients ($\alpha = \epsilon\ C$ and $\alpha^{'} = \sigma\ N$) are different, but can be converted using $\alpha^{'} = \alpha$ln(10).

$$A = ln(I_L^o/I_L) = \sigma N l \tag{2.53}$$

It would be logical to report CD data in terms of $\delta\epsilon$ (allowing for samples of differing concentration and path length to be compared directly) as modern instruments are capable of measuring small δA. However, another method of expression relies on the ellipticity, θ (Eq. 2.54), which is derived from the minor (b) and major (a) axis of the resultant elliptically polarised light (Fig. 2.4). Furthermore, these two quantities (i.e. δA and θ) are linked by a simple numerical relationship, θ = 32.98δA [1, 39, 40].

$$\theta = \tan^{-1}\frac{b}{a} \tag{2.54}$$

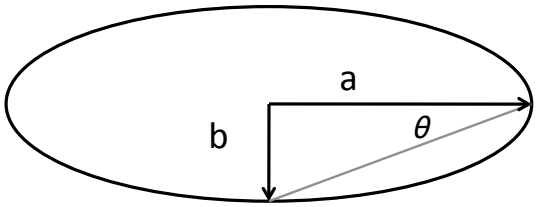


Figure 2.4: Diagrammatic representation of the elliptically polarised light, revealing the major axis, a, the minor axis, b, and the ellipticity, θ.

Spectra recorded in the near-UV and visible region are commonly expressed in terms of protein molarity, however in far-UV CD of polypeptides and proteins it is common to use the mean residue concentration. This is obtained by dividing the concentration in mass terms (i.e.

$mg.ml^{-1}$) by the mean residue weight (MRW), M_{mrw}, which is the molecular mass divided by $N-1$ (where N is the number of amino acids in the peptide chain). Far-UV CD spectra are therefore often expressed in terms of the "mean residue ellipticity" $[\theta]_{mrw}$ (Eq. 2.55) which is quoted in units of $deg.cm^2.dmol^{-1}$, and is related to the observed ellipticity (θ), the path length (d) in cm, and the concentration (c) in $g.ml^{-1}$.

$$[\theta]_{mrw} = \frac{\theta.M_{mrw}}{10.c.l} \tag{2.55}$$

Furthermore, one can express the concentration in terms of $mol.l^{-1}$ (C_M), and remove the dependence of M_{mrw} from the equation; the molecular mass is factored into the number of moles (g/MW), therefore one need only consider the number of amino acids, N (Eq. 2.56).

$$[\theta]_{mrw} \propto \frac{\theta}{C_M.l.(N-1)} \tag{2.56}$$

However as mentioned above, CD spectra may be expressed in terms of absorbance. Here the units relate to the "difference in molar absorbance" $\delta\epsilon$ and are expressed as $cm^{-1}.M^{-1}$. Like the observable quantities δA and θ, these two units $\delta\epsilon$ and $[\theta]_{mrw}$ are linked by a simple numerical relationship $[\theta]_{mrw}$ = 3298$\delta\epsilon$ at all wavelengths [1, 39, 40]. At this point it should be noted that henceforth all reference to the "mean residue ellipticity" will omit the 'mrw' subscript (i.e. $[\theta]_{mrw} \leftrightarrow [\theta]$).

2.2.1.3 Identifying Secondary Structure

The α-helix, β-sheet and random coil structures each give rise to a characteristic shape & magnitude of CD spectrum (Fig. 2.5). The α-helix displays large CD bands with negative ellipticity at 222 & 208nm and a CD band with positive ellipticity at 192nm. Similarly to the α-helix, the 3_{10}-helix displays CD bands with negative ellipticity at 222 & 207nm (the former being a shoulder), with a weak positive band near 195nm. The ratio of $[\theta]_{222}$ / $[\theta]_{207}$ is found to be much weaker than that of the α-helix. Overall the CD spectrum of the 3_{10} is different but NOT dramatically from those of α-helices.

The left-handed polyproline helix, pPII, gives a CD profile that is very similar to that of an unordered conformation. Both are known to give a CD band with negative ellipticity around

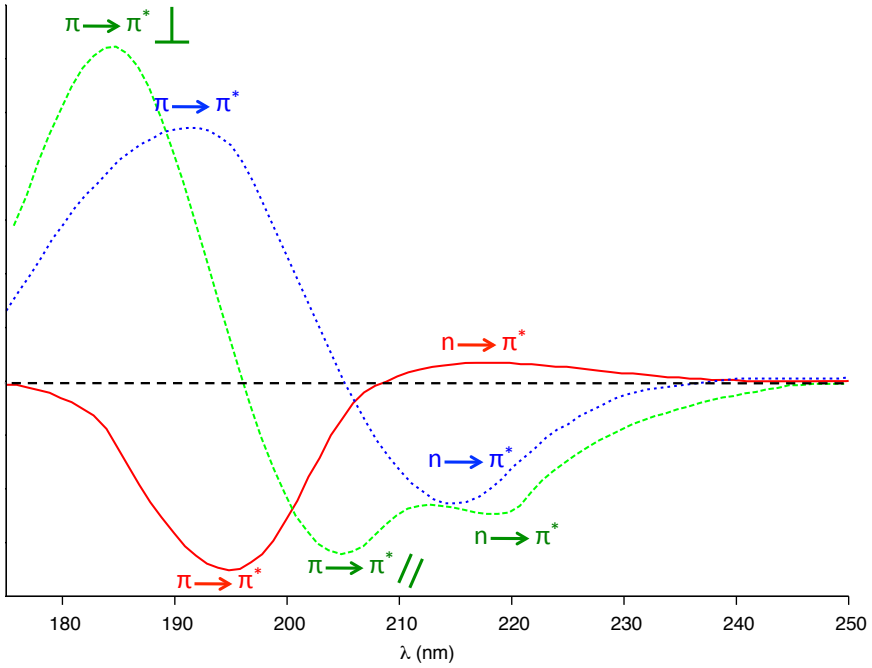


Figure 2.5: Example of far-UV CD spectra for an α-helix (green), β-sheet (blue) and random coil (red). Revealing the associated electronic transitions of the absorption bands: the exciton coupling of $\pi \rightarrow \pi^*$ results in two bands polarised perpendicular to, or along the helix axis (parallel) [1]

200nm, while the pPII helix is characterised by a weaker positive band at about 217nm. This positive band is found to shift in position to 225nm when the peptide is rich in proline.

β-sheets exhibit a broad negative band near 217nm and a large CD band with positive ellipticity near 195nm. The exact position of these two bands and their ratio are found to be more variable than those of the α-helix. This may be due to β-sheets being able to adopt a parallel or anti-parallel (or mixed) arrangement. Their ability to be twisted may also contribute to the variability in the bands features. In the case of β-turns, there is no CD spectrum that is characteristic of this structure. Type II β-turns exhibit a weak CD band with negative ellipticity at 220-230nm, with a stronger positive band between 200-210nm. Type I & III β-turns on the other hand give CD spectrum which are qualitatively similar to that of an α-helix. The distinction is in the two longer λ bands of negative ellipticity, with the amplitude of these bands for turns being approximately half of their α-helix counterparts. In addition the positive band near 190nm is $\frac{1}{4}$ to $\frac{1}{3}$ of the magnitude of an α-helix [1, 41–45].

2.2.1.4 Origin of bands

The OCN peptide bonding system consists of the single C-O and C-N σ bonds and a delocalised π bonding system (Fig. 2.6a) comprised of the p_x atomic orbitals and contains two electrons from the nitrogen, one from the oxygen and one from the carbon. Within the peptide bond two electronic transitions may occur. The first is the $\pi \rightarrow \pi^*$ transition from within this π bonding system. The oxygen atom also posses two nonbonding electrons in the atomic p_y orbital (Fig. 2.6b), and one of these may be excited to the π^* orbital - resulting in the $n \rightarrow \pi^*$ transition.

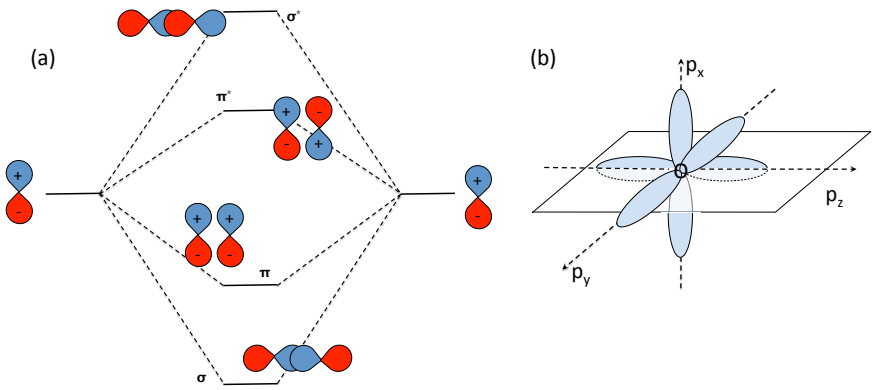


Figure 2.6: Molecular orbital energy level diagram from p-orbitals; illustrating end-on p-orbital overlap resulting in σ_p orbitals and side-on p-orbital overlap resulting in π_p orbitals. (b) The three p-orbitals of oxygen are located along the x, y, and z axis.

The $n \rightarrow \pi^*$ transition is of lower energy and manifests as absorption at longer wavelengths than the $\pi \rightarrow \pi^*$ transition (Fig. 2.5). In the case of the α-helix, $\pi \rightarrow \pi^*$ transition is split into two transitions as a result of 'exciton coupling' [1]. This occurs when one has two identical chromophores, one excited and the other not. Resonance of the excitation could occur between the two chromophores, and at any given time, one would be unclear as to which was in the excited state. This exchange results in two states, one of higher energy, and one of lower energy in regards to the two isolated (non-interacting) chromophores.

2.2.1.5 CD as a quantitative technique

Due to CD being a quantitative technique, it is highly regarded in following denaturation of proteins. CD obeys Beer's Law (Eq. 2.51) and therefore the spectrum of each component in a solution is directly proportional to its concentration. The total CD spectrum is essentially a sum

of the contributing spectra (Eq. 2.57).

$$\theta_{x(A+B)} = \frac{\theta_{x(A)} + \theta_{x(B)}}{2} \tag{2.57}$$

A further consequence is that a change in the CD spectrum due to addition of a denaturant or heat, is directly proportional to the concentration of native & denatured forms. As such, following these changes as a function of temperature or denaturant is a very convenient and easy way of obtaining information on the thermodynamics of folding.

2.2.2 Fourier Transform Infrared Spectroscopy

Fourier transform infrared (FTIR) spectroscopy has been widely used as a basis for secondary structure analysis [46]. The torsional angles, which determine the spatial orientation of the peptide backbone, ultimately affect the vibration modes of the peptide bond [47] - known as the amide modes. The understanding of these amide modes stems from early work on the simplest model compound which includes the amide group, N-methyl acetamide (NMA) [48]. From this, three classical amide modes have been identified. The amide I band is primarily the stretch of the C=O bond [7] and can be found in the 1610-1700 cm^{-1} region depending on the solvent. As it depends on the secondary structure of the backbone and is hardly affected by the nature of the side-chain, it is the most commonly investigated band in structural analysis [49]. Amide II, found in the 1530-1580 cm^{-1} region [50], is regarded as the out of phase combination of the N-H in plane bend and the C-N stretch [7, 51], and is considered less structurally sensitive than amide I. Finally, the amide III mode is the in phase combination of the N-H in plane bend and C-N stretch, and normally appears in the 1200-1340 cm^{-1} region [50].

Overall information on the secondary structure of peptides can be derived from analysis of the amide I band which is strong in IR [52]. In solution, the frequency of the C=O stretch is sensitive to hydrogen bonding characteristic conformation forms [53]. However, this band is not a pure band but consists of several components each associated with various secondary structures.

2.2.2.1 Identifying Secondary Structure

In IR analysis, experimental studies show that α-helical conformations are characterised by IR absorptions in the 1650-1658 cm^{-1} range [54]. However, studies have drawn attention to proteins which absorb in this region yet exhibit no α-helical structure [55–57]; large looped structures have been recorded to present bands around 1655 cm^{-1} [56, 57]. Another complication in the assignment of α-helices is the small number of proteins which absorb below 1650 cm^{-1}, yet are known to be predominately α-helical [58, 59]; one example is the homopolypeptide poly-L-lysine, which absorbs at 1638 cm^{-1}.

Other helical forms such as the 3_{10}-helix are also difficult to assign on the basis of IR absorption. Synthetic α-aminobutyric acid-containing peptides, known to form 3_{10}-helices, have been shown to characterised by strong amide I bands at 1662-1663 cm^{-1} [60]. Whereas a band centred near 1640 cm^{-1} has been assigned to the repeating i, $i+3$ hydrogen bonds of a 3_{10}-helix [61, 62]. The type III β-turn, regarded as the building block of a 3_{10}-helix, has also been predicted to absorb in this region [60]. In fact FTIR studies conducted on i, $i+3$ hydrogen bonds in both linear and cyclic peptides, suggest that all three of the main β-turn subtypes (i.e. I, II, III) present characteristic absorptions near or below this region. Bands occurring in the high frequency amide I region (>1660 cm^{-1}) have also been associated with turns [63].

IR absorption in the 1620-1640 cm^{-1} region corresponds to pleated β-sheets [52, 53]. An antiparallel β-sheet structure is characterised by a strong band $\sim$1632 cm^{-1} irrespective of the medium in which the spectrum is recorded, and a (weaker) high wavenumber component located at 1690 cm^{-1} in H_2O or 1675 cm^{-1} in D_2O [7]. It is often claimed that a distinction can be made between antiparallel and parallel sheets due to the theoretical prediction that an infinite parallel sheet should lack the higher wavenumber component [64]. Experimentally however, there is no evidence for this claim and similar spectra are expected for finite β-sheets [64–67]. Therefore contrary to results suggested by theoretical calculations, a distinction between antiparallel and parallel sheets does not appear possible on the basis of the amide I band [65, 66]. The β-hairpin bands are seen to differ in two aspects from ordinary antiparallel β-sheets, with a low wavenumber absorption at 1620 cm^{-1} and a higher wavenumber absorption that does not vary between H_2O and D_2O [68]. Finally, absorption in the region 1640-1648 cm^{-1} is commonly assigned to backbone structures which are unordered or aperiodic [53].

2.2.3 Dynamic Light Scattering

Dynamic Light Scattering (DLS) is a useful technique which measures the Brownian motion of a particle and relates this to the particles size. Brownian motion itself refers to the random movement of a particle caused by collisions with the surrounding solvent molecules. These collisions result in smaller particles (with less inertia) being propelled further and faster than larger particles. The velocity of this Brownian motion is defined by the translational diffusion coefficient, D, which can be used to determine the size of the particle via the Stokes-Einstein equation (Eq. 2.58)

$$d(H) = \frac{kT}{3\pi\eta D} \tag{2.58}$$

The size that is actually computed is the hydrodynamic diameter, $d(H)$, and refers to how the particle diffuses within a liquid. This value equates to the diameter of a sphere which has the same translational diffusion coefficient as the particle. It should be noted that as the viscosity, η, of a liquid relates to its temperature, T, DLS requires that one accurately knows the temperature and that it must be stable, otherwise convection currents in the sample will result in inaccurate results due to non-random movements.

The speed at which the particles are diffusing is evaluated by measuring the rate at which the intensity of scattered light fluctuates. Theoretically, if one was to use a laser to illuminate a cuvette filled with stationary particles, one would observe (via a frosted glass screen) a classical speckle pattern consisting of patches of bright light and dark areas. These speckles arise from the scattering of the propagated waves of light by the particles (Fig. 2.7). The bright patches are caused by light waves arriving with the same phase at the detector and constructively interfering, whereas the dark areas are due to phase additions that result in destructive interference. In this theoretical example, as the particles are stationary, the sizes and positions of the speckle pattern will also be stationary. For a system undergoing Brownian motion, the corresponding speckle pattern will be in constant motion with the rate of fluctuations in the intensity being dependent on the size of the particles.

A digital correlator is then used to measure the degree of similarity between the intensity signal at a particular time, t, and at a very small time later, $t+\delta t$, and it would be expected that these

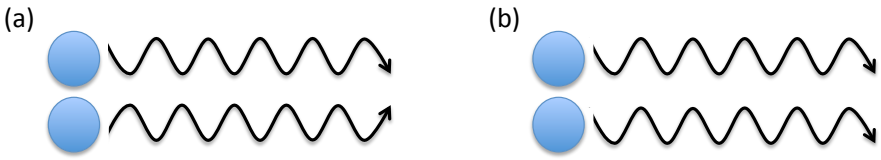


Figure 2.7: The observed signal is dependent on the phase addition of the scattered light, (a) decreased intensity is a result of two beams interfere and cancel each other, (b) while increased intensity results if two beams interfere and enhance each other.

values would be very similar (strongly correlated). However as one compares values at greater separations (i.e. $t + 2\delta t$ etc.) this correlation would begin to reduce due to the random nature of Brownian motion (Fig. 2.8a). Assigning perfect correlation as being 1 and no correlation as 0, one can plot the correlation as a function of time (Fig. 2.8b).

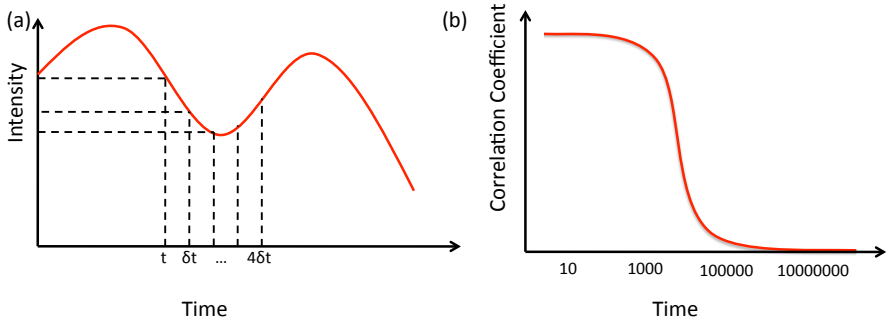


Figure 2.8: (a) The intensity of scattered light may fluctuate as a function of time. (b) Following the correlation with respect to time, one would expect the correlation of larger particles to decay

The rate at which this correlation decays is linked to the size of the particles. As smaller particles move more rapidly due to Brownian motion, their intensities will also fluctuate more rapidly, resulting in their corresponding correlation functions decaying faster than larger particles.

2.2.4 Fluorescence Spectroscopy

Fluorescence is one of the two categories of luminescence, which is defined as the emission of light from any substance; the other category is termed phosphorescence, and is distinguished from fluorescence in terms of the nature of the excited state from which the light is emitted from (fluorescence, singlet; phosphorescence, triplet). The major difference between the two forms of luminescence, is that fluorescence occurs between two singlet states, an allowed transition, while

phosphorescence occurs between a triplet and singlet state, a forbidden transition (Fig. 2.9). This results in emission rates for phosphorescence being significantly slower in comparison to those of fluorescence .

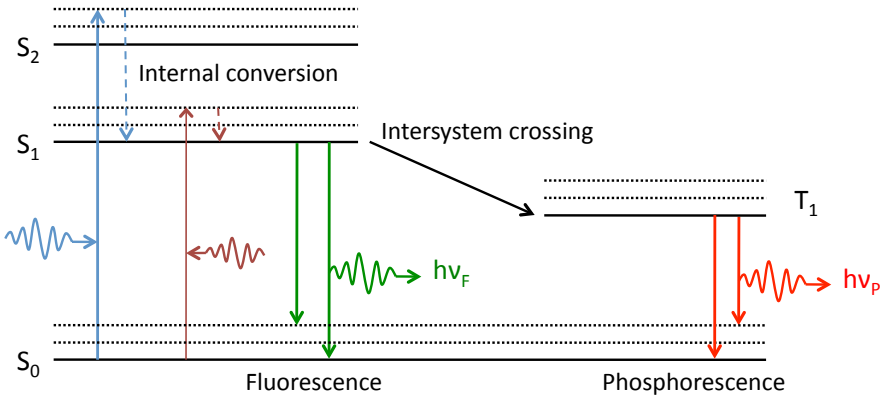


Figure 2.9: A typical Jablonski diagram, highlighting the difference in fluorescence and phosphorescence. The singlet ground, first and second excited states are denoted S_0, S_1 and S_2 respectively, while the first triplet state is denoted T_1.

Typically, emission from fluorophores occurs at longer wavelengths than those of absorption, due to a variety of dynamic process that transpire after light absorption (Fig. 2.10). For example, when a fluorophore is excited to the first singlet state (S_1), it is usually to an excited vibrational level within S_1, and the excess vibrational energy is rapidly lost to the solvent. Additionally, if the fluorophore is excited to the second singlet state (S_2), internal conversion results in it rapidly decaying to S_1 within 10^{-12}s. Within the excited state (μ_E) the fluorophore generally has a larger dipole moment than the ground state (μ_G), thus the solvent dipoles can reorient around μ_E and lower the energy of the excited state; the greater the polarity of the solvent, the greater the effect. Naturally, non-polar molecules (e.g. aromatic hydrocarbons) exhibit less sensitivity to solvent polarity than polar fluorophores [69].

Among biopolymers, proteins are unique in that they display intrinsic fluorescence; proteins may contain three amino acids which are all fluorescent - phenylalanine, tyrosine, and tryptophan. As these residues are typically rare in proteins, it allows one to interpret the spectral data. Tryptophan itself is highly sensitive to its environment, with changes in its emission spectra resulting from conformational transitions, subunit association, denaturation or substrate binding. Within a protein containing multiple tryptophans, while each residue may be in a unique envir-

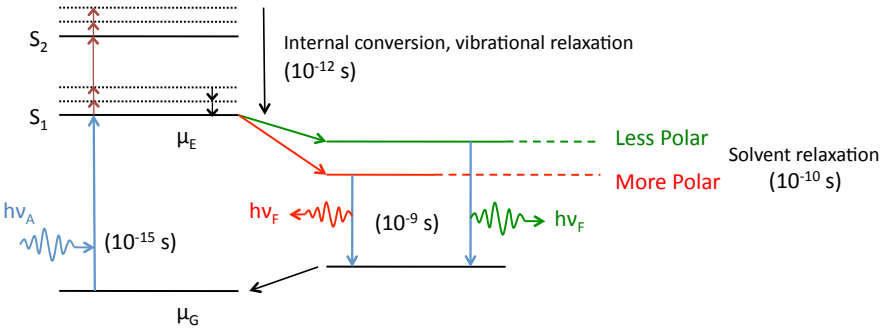


Figure 2.10: Jablonski diagram revealing the effects of solvent relaxation.

onment, the absorption and emission spectra of tryptophans typically overlap and it is difficult to separate the spectral contributions of each residue [70].

An important feature for this research is that the tryptophan emission is dependent on the structure of the protein. Within an apolar environment, one would expect the tryptophan emission to be blue-shifted due to the aforementioned solvent relaxation [69, 71]. As the tryptophan experiences hydrogen bonding, the emission shifts towards longer wavelengths (Fig. 2.11). This is analogous to the tryptophans becoming solvent exposed, and denaturation of proteins results in similar emission spectra. Similarly, if the emission spectra is blue-shifted, then one may deduce that the tryptophans have become less solvent exposed and are in a hydrophobic environment.

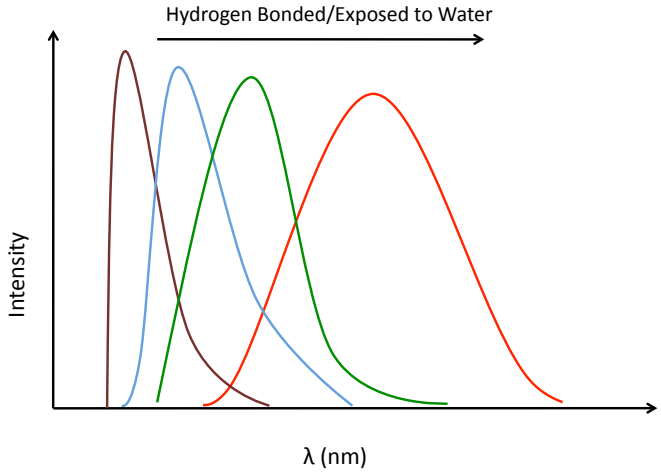
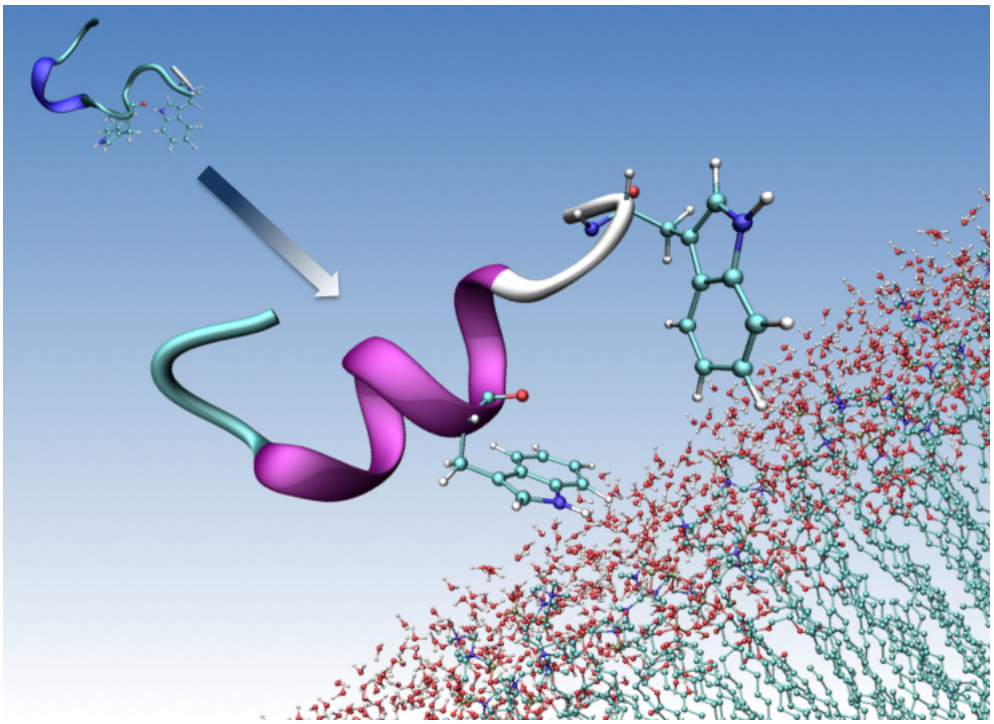


Figure 2.11: Effect of the tryptophan environment on the emission spectra.

3

Environmental bias: Conformational equilibria of $gp41_{659-671}$.



3.1 Rationale

The structure of certain peptides can be regarded as being comprised of an ensemble of flexible conformations, characterised by a rugged energy landscape with shallow minima. Such a peptide should be sensitive to changes in the environment (i.e. solvent), which may favour the equilibrium of structures towards a particular population (or fractional population). It is well documented that the conformation of proteins and peptides are greatly influenced by interactions with different solvent environments [72–74] and a solvent can override the propensity for a secondary structure.

Membrane surfaces have been shown to induce a partial denaturation of proteins resulting in a non-native state [75, 76]. This effect has been proposed to be due to a cumulative effect of the negative charges of a membrane [77] and a decrease in dielectric constant near the membrane surface [78–80]. The conformational changes experienced by a membrane active peptide under such solvent influences have been shown to result in structure similar to that observed upon interaction with membranes [72, 73].

The effect a membrane has on a protein's structure is dependent partly on the location of the protein relative to the membrane, due to various environmental conditions that it may experience [81]. These include diffuse layers where the dielectric constant can be either considered well approximated by aqueous conditions (at far distances) or is considerably reduced (at intermediate regions of closer approach); and the interfacial contact regime where the membrane is in contact or in very close proximity to the protein.

3.2 System chosen

A major problem in combating HIV is that the virus has evolved numerous and sophisticated immune-evasion strategies, many of which rely on multiple mutations and sequence variability [82]. The Membrane Proximal Region (MPR) of the HIV-1 fusogenic subunit (gp41) is found to be a highly conserved sequence, and therefore an attractive focus for new therapies. Of the few broadly neutralising monoclonal antibodies (mAb) which have been isolated [83], three target gp41 (2F5, Z13 & 4E10) [84, 85] with epitopes localised in the same continuous MPR; this region encompasses amino acids 660-683 (numbering based on HIV-1 HxB2 [86]) and also

includes the C-terminal residues relevant for the anti HIV-1 fusion inhibitor T-20 (Fig. 3.1a).

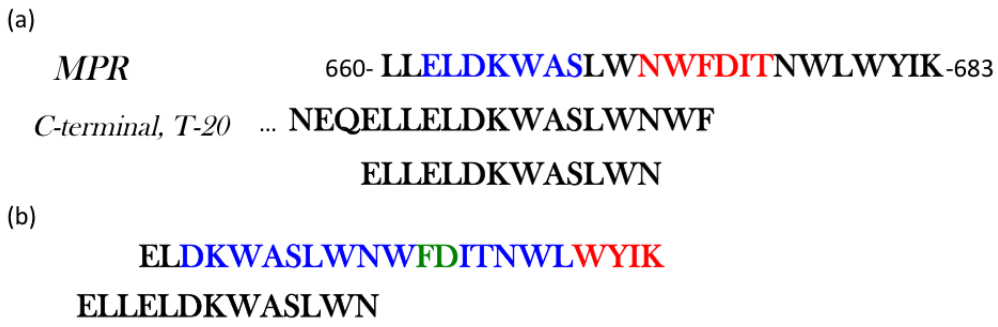


Figure 3.1: (a) The full sequence of the membrane proximal region (MPR) $gp41_{660-683}$ highlighting the epitopes for two broadly neutralising antibodies, 2F5 (blue) and 4E10 (red). The C-terminal sequence is also shown to illustrate the amino acids in common with the peptide ELLELDKWASLWN, the object of this study. (b) The Structure of the peptide ELDKWASLWNWFDITNWLWYIK was solved by NMR in the presence of DPC micelles. The conformation adopted is of a kinked α-helix. Blue: amino acids in α-helix, Red: 3_{10}-helix.

Approximately half of the residues in this region are hydrophobic with five of these being tryptophan [87]. Mutations in this region (in particular the Trp residues) have been shown to decrease the fusogenicity and inhibit viral entry, suggesting that these residues are critical for the function of this domain [88]. However attempts to elicit a neutralising immune response from the membrane-proximal region, and especially the hexapeptide epitope ELDKWA ($gp41_{662-667}$) [2] of the human monoclonal antibody (mAb) 2F5, have not succeeded [89, 90]. This has prompted several studies to determine the structure of this sequence as well as to establish whether or not additional residues outside of the 'core' hexapeptide are required for successful binding to the 2F5 mAb [2]. It has also been reported the 2F5 epitope must contain the sequence $gp41_{659-671}$, where the additional residues enhance the binding affinity by a factor of 10^3. It is this sequence that we focus on in this study.

Previous solution studies of $gp41_{659-671}$ have led to conflicting findings regarding conformational preference [2, 91]. Nuclear Magnetic Resonance data implied that this peptide has a strong conformational bias toward the rare 3_{10}-helix in water. Conversely, a greater variety of local motifs in this peptide has been revealed by UV resonance Raman scattering [92] and independent circular dichroism data [2, 93] revealed no strong helical bias under aqueous conditions. However, parallel tempering molecular dynamics simulations using the CHARMM empirical potential were strongly biased toward helical motifs in water accounting for the experimental

data only at unphysically high temperatures [93].

Structural data has shown that the sequence gp41$_{662-683}$ forms a kinked helix in liposomes with a hinge centred on gp41$_{673-674}$ (Fig. 3.1b) [94, 95], which is in contrast to previous results pertaining to the shorter sequence gp41$_{665-683}$ which revealed a straight α-helix [96]. The functional role of these two different conformations has been proposed and linked to mAb neutralising activity; suggesting that the interconversion between the two conformations is an important aspect of the fusion process. Furthermore, it has been proposed that the capping residue Asn$_{671}$ acts as a switch in this interconversion between the kinked conformation and the longer α-helix [97]. These studies have also demonstrated that both the nature of the lipid and the pH were important factors in the structure of the P1 peptide (residues 649-683), located in the conserved MPR of gp41 [94]. At neutral pH, the N-terminal charged residues of this sequence adopt a disordered structure, while the Trp-rich domain folds into a helix in order to interact with DPC micelles. Under acidic conditions, the Glu-rich region of the P1 peptide also folds into this helical conformation.

We propose that if the hinge gp41$_{673-674}$ is indeed functionally important, then the sequence gp41$_{659-671}$ should exhibit an independent capability to interact with a membrane surface. It is also proposed that the two Trp residues are able to fold independently in order to interact with the membrane surface. This autonomous interaction with the surface of a membrane is supported by an investigation into interactions of dipeptides with SDS micelles which probed the effect of neighbouring amino acids on the energetics of the interactions [98], concluding that the presence of either lysine or leucine preceding a Trp residue greatly favoured the Trp-micelle interaction. As both of these amino acids are present in the sequence gp41$_{659-671}$ (ELLELDKWASLWN), this therefore implies that this peptide should posses a strong affinity for binding. Finally, as these Trp residues may be assigned positions i and $i + 4$ relative to one another, it is logical to assume that these fold into an α-helix.

Here we examine the secondary structure of gp41$_{659-671}$ using a combination of computer simulation and experimental measurements under various solvent conditions. These are chosen in order to mimic the distinct physical and chemical environments that this sequence may be exposed to during the viral fusion process. The aim of which is to uncover aspects of structural control and membrane binding of the peptide by chemical and physical processes accessible in

the extracellular environment.

3.3 Materials & Methods

3.3.1 Materials

The amino acid sequence, ELLELDKWASLWN, was purchased from Cambridge Peptides with a stated final purity of 100%. The peptide stock was prepared by dissolving the peptide powder directly in 20mM buffer solution when applicable. For samples in pure water, the solubilisation of the peptide was helped through additions of 2μl of 1M NaOH. In order to check for complete dissolution, the absorbance at 350 nm was monitored. The concentration of the stock solutions were determined spectrophotometrically using the calculated extinction coefficient at 280 nm of 14000 M^{-1} cm^{-1} [91] due to the presence of two tryptophan residues in the peptide sequence.

The peptide stock solutions were used to prepare working solutions of various concentrations of Trifluoroethanol (10%, 20%, 30% v/v), Methanol (50% v/v), Sodium Dodecyl Sulphate (3mM, 4mM, 5mM, 6mM, 9mM, 12mM), and Acetonitrile (50% v/v). Water and 0.2M buffer solutions were used to ensure the overall buffer concentration remained 20mM. The concentration of each sample was checked spectrophotometrically prior to any experiments being carried out, using the same extinction coefficient used to determine the concentration of the stock solution.

3.3.2 Circular Dichroism (CD) measurements

Far-UV CD spectra were recorded on a Jasco J-810 spectropolarimeter (Japan Spectroscopic Co., Tokyo) fitted with a Peltier unit for temperature control. The concentration of peptide in solution was 0.04mM. The spectra are the average of three accumulations in step mode (data pitch of 1 nm, response time of 8s, bandwidth of 1 nm) acquired in rectangular cuvettes of 0.1 cm path length (unless otherwise stated) at 25°C. For every sample, an appropriate solvent spectrum was recorded under identical conditions and subtracted from the corresponding sample spectrum. The data collected was expressed in molar ellipticity (deg.cm^2.$dmol^{-1}$) in order for a direct comparison with CD spectra reported to date [2, 91].

3.3.3 Dynamic Light Scattering measurements

DLS was used to confirm whether the concentrations of the surfactants were above or below the critical micelle concentration (cmc). The DLS measurements were performed using a Zetasizer Nano (Malvern) and disposable clear plastic cuvettes at 25°C. For each run, three accumulations were performed to ensure consistent results. Data was subsequently analysed using the manufacturer's Dispersion Technology Software (DTS version 5.10).

In order to ensure subsequent interpretation of the CD spectra was handled correctly, it was necessary to confirm which of the concentrations of SDS lay above and below the cmc. DLS was performed on the solutions of various concentrations of SDS in 20mM sodium phosphate (i.e. without $gp41_{659-671}$) to determine where they lay relative to the cmc of the surfactant (Fig. 3.2a). Furthermore, in order to verify that the solutions of $gp41_{659-671}$ and SDS under the different pH conditions tested lay above the cmc, DLS measurements performed on the samples (Fig. 3.2b) revealed that all samples lay above the cmc.

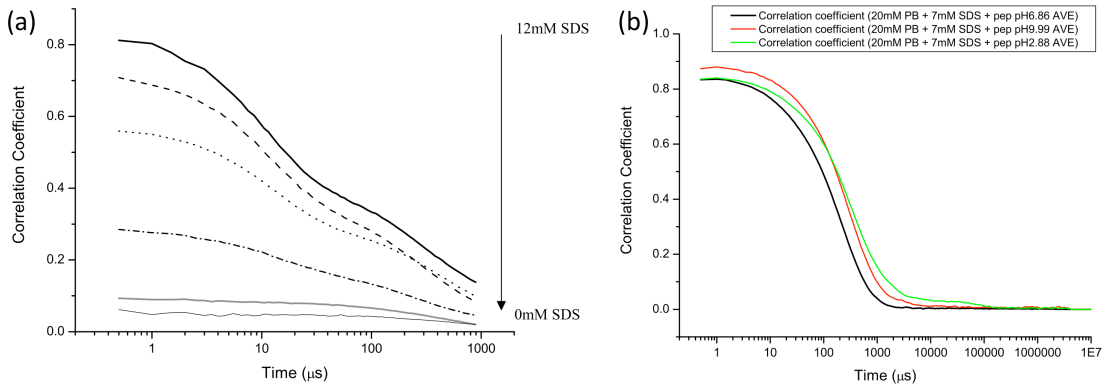


Figure 3.2: (a) Correlograms of solutions at different SDS concentrations. (b) Correlogram of a solution of 7mM SDS in 20mM PB at different pH: demonstrating that under all conditions they were above the cmc.

The nature of the correlograms belonging to the various concentrations of SDS would indicate the presence of polydisperse aggregates, as evident by the multiple exponential decays (Fig. 3.2a). Analysis of the hydrodynamic radius for the samples at different concentrations of SDS revealed the emergence of micelles of 4.187 - 4.85nm above concentrations of 4mM; the length of an SDS monomer is approximately 17Å. Traces of larger aggregates (100-200nm) were also detected utilising the manufacturers software; the intensity of the 4nm species was found to become the most prominent of the two, as the concentration of SDS increased.

Analysis of the samples of gp41 in 7mM SDS revealed average hydrodynamic radii of 118.3nm, 192.8nm and 266.8nm for the solutions near pH 7, 10 and 3 respectively; decorrelation times of 0.193ms, 0.304ms, and 0.369ms, derived from the correlograms, corresponded to sizes of 24.2nm, 38.1nm and 46.2nm. This suggested the presence of larger aggregates potentially comprised of a mixture of SDS and peptide.

3.3.4 Fluorescence Spectroscopy

Tryptophan fluorescence was measured with a LS55 (Perkin Elmer Instruments, Massachusetts, USA) at 25°C using a 1cm quartz cuvette (Hellma GmbH, Germany) and 0.0647 mg ml^{-1} protein solution. To selectively excite the tryptophan an excitation wavelength of 295 nm (2.5nm excitation slit) was used. The spectra were acquired from 300 to 400 nm (emission slit 6nm) and subtracted with the corresponding buffer blank.

3.3.5 Computational Methods

Three starting configurations based on motifs reported by Barbato *et al.* [2] and Biron *et al.* [91] have previously been investigated using classical atomistic molecular dynamics [93]. One motif that was reported by Barbato, yet not investigated, was that of the unordered conformation. Furthermore, as the results of the previous study suggested that the CHARMM force field over emphasised the helical content in pure water, it was imperative to investigate the effect of employing different force fields for this system. A unit cell was set up, which comprised one gp41 molecule (Fig. 3.3), 1853 water molecules and 2 sodium counter ions in order to ensure charge neutrality. The molecules were arrayed on a cubic lattice with random molecular orientations in a cubic box of length 38.2Å. The water molecules were represented by the rigid, non-polarisable 3-site TIP3P empirical force field. The unit cell was replicated three-fold with the peptide being described by either the CHARMM22 [31], AMBER-ff03 [36, 37] or AMBER-ff99SB [35] force fields in each of the replicas, while the PINY simulation package [99] was employed to perform the simulations.

The systems were equilibrated at 300K in the canonical ensemble for 150ps with a time-step of 0.5fs to anneal out unphysical contacts. The system was then run for 150ps with a time step of 0.5fs in the isothermal-isobaric ensemble to allow for spatial relaxation. The system was

then run for a further 150ps in the canonical ensemble with a time-step of 1fs being employed, before a production run of 15ns using a time-step of 2.5fs from which all data was collected. Periodic boundary conditions were employed and long-range interactions were evaluated via Ewald summation.

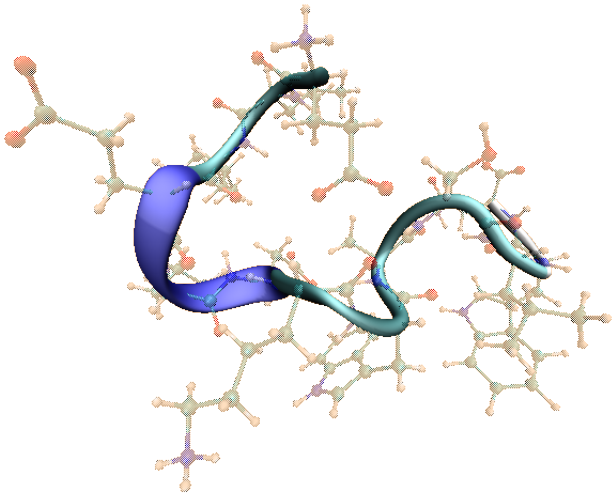


Figure 3.3: Initial starting configuration used in simulations, as reported by Barbato *et al.* [2]

In order to analyse the structure of the peptide using Molecular Dynamics simulations, we have made use of Ramachandran plots and bond probability distribution functions, as used by Samuelson *et al.* [100]. The former allows for the assignment of secondary structure motifs based on dihedral angles, while providing a clear illustration of the conformational space explored during the course of the simulation. The later on the other hand allows for the probability distribution of selected atomic contacts as a function of inter-atomic separation to be assessed. These distribution functions arise from the connectivities between residues i and $i + n$ where n=3, 4 or 5 for 3_{10}-, α- and π-helices respectively, and examination of such functions allows us to assess the conformational and secondary structure preferences for each individual residue at a given temperature.

3.4 Results

3.4.1 Structure under aqueous conditions

3.4.1.1 Structure obtained experimentally

Due to contrasting results obtained in literature [2, 91], far-UV CD spectra of 0.04mM $gp41_{659-671}$ in various buffers (water, 20mM ammonium acetate, 20mM sodium phosphate) were recorded. The buffers were chosen to match those employed in these past studies, in order to determine if the choice of buffer ultimately affected the secondary structure of $gp41_{659-671}$. No discernible structural change was observed between the conditions tested (Fig. 3.4), confirming that the environment of the buffer was not responsible for these reported discrepancies. The overall shape of the spectra were identical to those reported previously by Barbato *et al.* [2], and were consistent with an aperiodic structure. This was interpreted from the location of the large band of negative ellipticity centred around 197 nm, both helical and β-sheet secondary structures are known to give rise to bands of positive ellipticity in this region. These two secondary structures also give large bands of negative ellipticity, however they are red shifted (relative to the location of our band) to 208 nm (plus 222 nm shoulder) and 217 nm respectively, thus ruling these structures out as a dominant secondary structure motif. CD spectra indicative of type II β-turns exhibit a weak band of negative ellipticity at 220-230 nm and a stronger positive band between 200-210 nm, whereas type 1 & III β-turns are known to give CD spectrum which are qualitatively similar to that of an α-helix. As these features were also not present in the spectra collected, these structures could also be ruled out as a dominant secondary structure motif.

It should be noted however that these structures should not be ruled out entirely. Barbato *et al.* [2] reported four conformations that collectively should be present in any feasible solution of the secondary structure. Theses conformations were that of an inverse γ- and type I β-turn, a helix with ϕ, ψ values intermediate between an α-and 3_{10}-helix, and finally an aperiodic strand. Additionally, reported molecular dynamics simulations of $gp41_{659-671}$ [101] revealed the secondary structure of the peptide varied significantly over the course of the simulation. Both of these studies suggest a dynamic equilibrium between different secondary structural elements with no dominant structure emerging.

Temperature dependence studies conducted by our group [93] revealed that the CD spectra

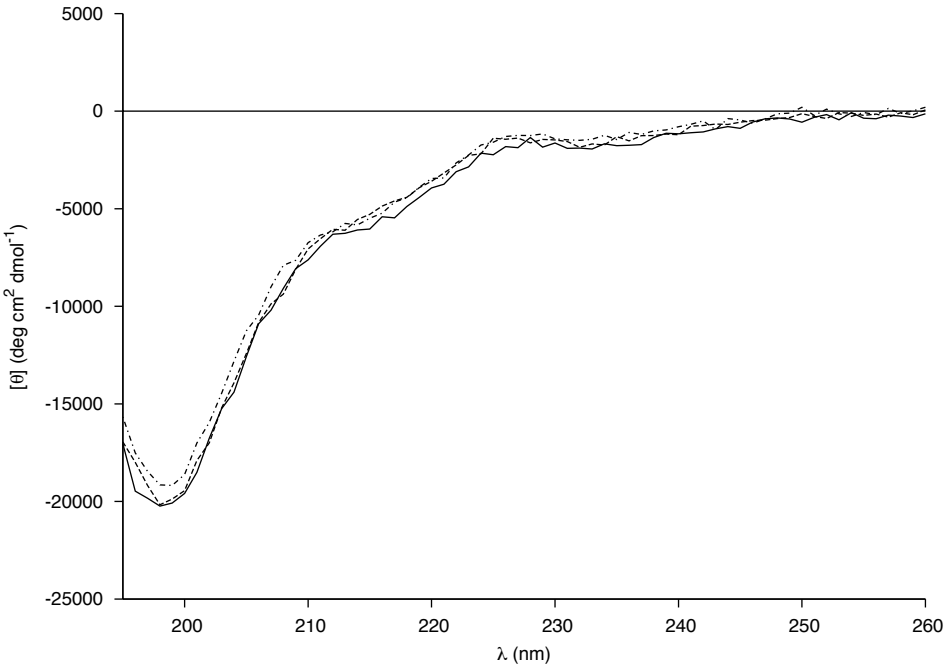


Figure 3.4: Far-UV CD spectra of 0.04mM gp41$_{659-671}$ in Acetate buffer pH 7.57, phosphate buffer pH 6.93 and H_2O pH 7.7.

of gp41$_{659-671}$ did not vary greatly at high temperatures when compared to those under ambient conditions. This indicated only a small change in the conformational population and pointed towards the structure of gp41$_{659-671}$ being consistent with a flexible conformational ensemble. It is well documented that under conditions thermodynamically favouring the native structure, protein structures are not static but in fact fluctuate on a rugged energy landscape that exhibit many local energy minima [102–104]. The CD spectra reported here are certainly consistent with a rugged energy landscape characterised by shallow minima. The individual amino acid residues along the sequence could adopt, transiently, ϕ and ψ angles typical of secondary structure elements, or there could be an equilibrium between different conformers at a chain level.

It has previously been proposed that this flexibility was important for the physiological function of this part of gp41 [2]. When an ensemble of conformers is present whether at residue or polypeptide chain level, one would expect the peptide to be sensitive to changes in the environment (i.e. solvent), which may favour the equilibrium of structures towards a particular

population (or fractional population). It is also possible that different elements of secondary structures, present at the sequence level, retain different degrees of stability relative to parameters such as solvent, temperature and pH. It is therefore of interest to investigate whether particular conformations can be promoted over other competing structures by altering such parameters.

3.4.1.2 Comparison to Simulation

Previous work has focused on investigating the conformational landscape of $gp41_{659-671}$ and the stability of previously reported structural motifs [93], by means of MD simulations employing disparate starting configurations; an α-helix, a type I β-turn, and an inverse γ-turn. This study made use of replica exchange (parallel tempering) in order to improve sampling of the rugged energy landscapes, a feat not easily achievable by conventional molecular dynamics simulations. Having identified the type I β-turn as being an unstable solution motif, attention was switched to the two remaining energetically competitive motifs. Furthermore, in order to investigate the competition between these two structures (i.e. α-helix and inverse γ-turn), an array of replicas was constructed in which the starting configurations rotated (e.g. ABABAB...) thus eliminating any bias introduced by the starting configuration. The structures encountered at 300K were predominantly helical (interestingly the initial structure at this temperature was the turn), and were in contrast to the flexible ensemble predicted experimentally (Fig. 3.4). In fact, the simulations were only found to be consistent with the data acquired by CD at temperatures in excess of 800K, at which the peptide was predicted to exhibit increased flexibility and disordered structure. In comparison to other MD studies [101, 105], the results obtained by means of replica exchange suggested that the CHARMM22 force field excessively stabilised helical structures at the expense of other structural elements. The force field therefore introduced an "effective temperature shift", in which an ensemble of secondary structure elements at residue level was only achievable at artificially high temperatures.

3.4.1.3 Need for New Force Field

In order to probe the chemical physics associated with structural variation near a membrane surface, it is desired to examine the secondary structure of $gp41_{659-671}$ in a mixed-solvent using long-time averaged molecular dynamics simulations. Unfortunately, the previous simulations

utilising the CHARMM22 force-field revealed that CHARMM22 over-stabilised the helical content when compared to experimental findings regarding the secondary structure of the peptide. A recent and thorough study suggested that classical force-fields in general over-express helical content by excessively stabilising helical content at the expense of other secondary structure motifs [32]. Furthermore, this study suggested that classical force fields may differ widely in the regions of dihedral space sampled. Before one can investigate the changes in secondary structure of $gp41_{659-671}$ in a mixed-solvent, it is necessary to find a force-field that yields a better agreement with the results obtained by far-UV CD. To that end, three separate force-fields were investigated in conjunction with a previously untested starting configuration reported by Barbato *et al.* [2] (i.e. unordered conformation). CHARMM22 was used in order to draw comparisons with other work using this force-field, but with disparate starting configurations. The additional two force fields were different versions of the AMBER force field, AMBER-ff03 and AMBER-ff99SB; both developed from the AMBER-ff99 force field in order to improve the balance between helical and extended regions of peptide and protein backbones via modifications to the ϕ and ψ backbone torsions.

Before analysing the secondary structure in detail, it should be noted that for the purposes of this work all references to the epitope will be in regards to the historic sequence $gp41_{662-667}$ (i.e. ELDKWA) rather than the extended sequence proposed by Barbato *et al.* [2].

Visualisation of the peptide after the initial equilibration revealed a structural change having occurred in each of the simulations (Fig. 3.5). Surprisingly, the two AMBER force fields appeared to display the greater helical content over the epitope.

Analysis of the backbone torsional angles over the initial 1ns for each of the simulations (Fig. 3.6) clearly showed that the epitope ELDKWA of the peptide was dominated by a helical secondary structure. In the case of the CHARMM22 model, it was clear that over the course of the initial 1ns, part of the epitope had additional secondary structure out with that of a helix. This was identified as belonging to the Asp_{664} residue. However the Ramachandran plots of the two AMBER force-fields seemed to indicate that the secondary structures of the amino acids out with the epitope were more stable over this initial 1ns, as the spread of the angles corresponding to each amino acid were more localised and not as diffuse.

Visualisation of the peptide after 10ns revealed further structural changes to the peptide in

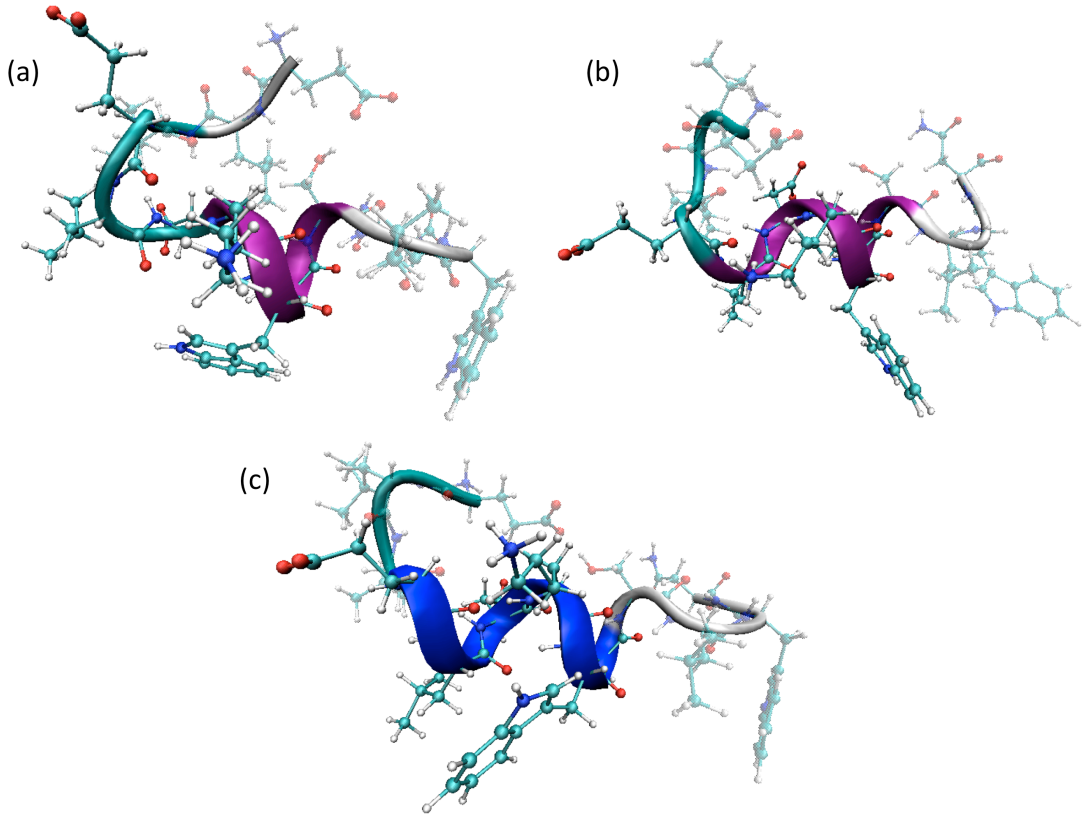


Figure 3.5: Secondary Structure of gp41 after equilibration: (a) CHARMM, (b) AMBER-ff03, (c) AMBER-ff99SB.

the simulations (Fig. 3.7). The simulation using the CHARMM22 force field had folded into an α-helix, spanning essentially the entire length of the peptide chain. The simulation employing the AMBER-ff03 force-field however clearly revealed that only the epitope displayed an α-helical motif, with residues lying out with the epitope adopting a more frayed-like structure. This structure was very similar to that observed after equilibration, indicating that this arrangement of different secondary structures along the peptide chain may be a stable arrangement for the AMBER-ff03 system. Visualisation of the simulation employing the AMBER-ff99SB force-field revealed an α-helical structure that extended beyond the epitope region yet not fully encompassing the whole peptide sequence.

Analysis of the backbone torsional angles over the last 5ns of each simulation (Fig. 3.8) revealed that in the system employing the CHARMM22 force field, the main feature of the Ramachandran plot of the epitope was a compact peak located within the right-handed α-helical region of dihedral space, with only residue Trp_{670} presenting a peak outside of this region (loc-

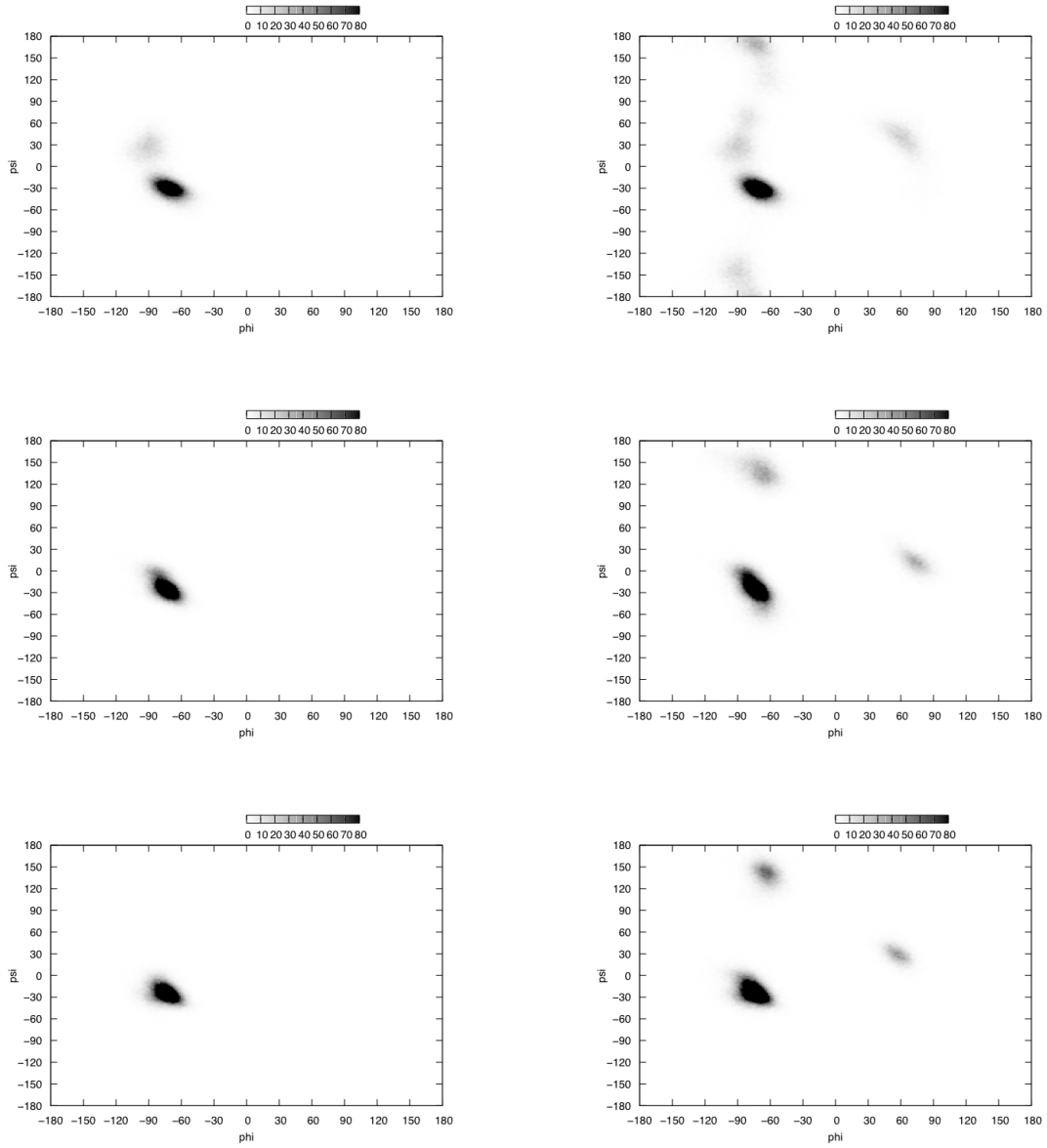


Figure 3.6: Ramachandran plots of the epitope (left) and full sequence (right) over the initial 1ns: CHARMM (top), AMBER-ff03 (middle), AMBER-ff99SB (bottom)

ated within the left-handed α-helical region). Under the AMBER-ff03 parameterisation, the main feature was also a compact peak located within the right-handed α-helical region, with the peak located within the β-sheet region arising from the torsional angles of the Leu_{661} and Leu_{669} residues, and the other peak arising from the Trp_{670} residue. Finally in the system defined by the AMBER-ff99SB force field, like the other two simulations, the main feature was a compact peak located within the right-handed α-helical region of dihedral space. The other two peaks out with

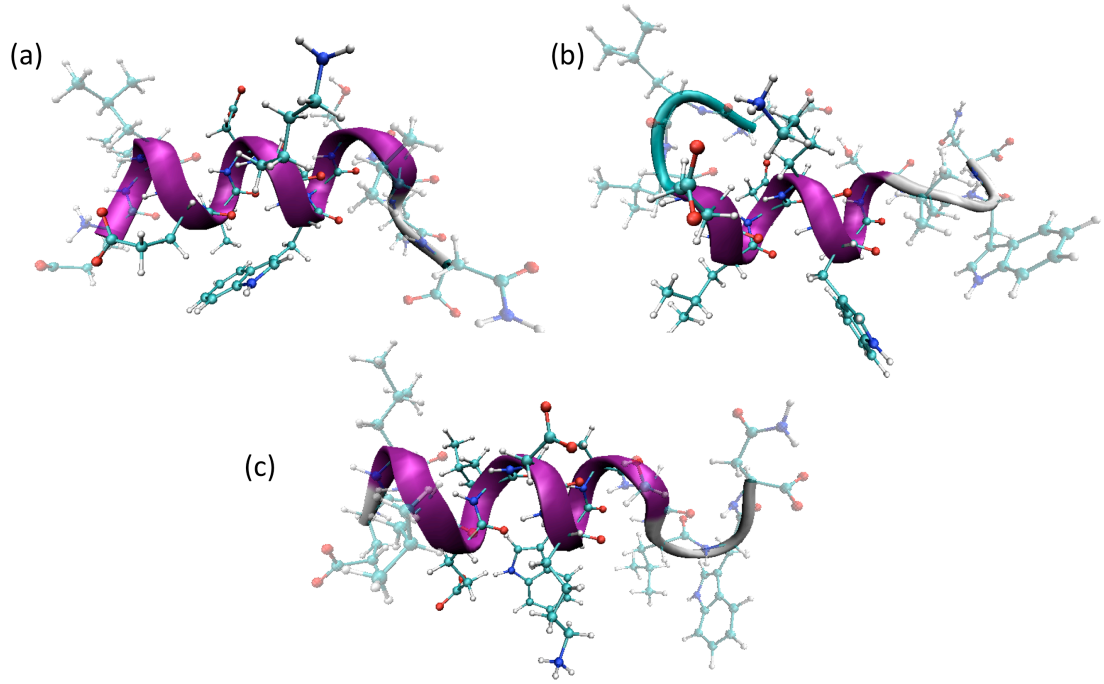


Figure 3.7: Secondary Structure of gp41 after 10ns: (a) CHARMM, (b) AMBER-ff03, (c) AMBER-ff99SB.

this α-helical region have been attributed to residues Leu_{669} and Trp_{670} for the peaks residing within β-sheet region and left-handed α-helical region respectively.

Bond probability plots offered a means to characterise the nature of the helical structure in more depth. Analysis of the bond probability distributions over the epitope for i, $i+3$ bonds corresponding to the Hydrogen-bonding in a 3_{10}-helix revealed that in the CHARMM22 model, residues Glu_{662}-Trp_{666} displayed peaks in the range 2.455 - 2.847Å (Fig. 3.9). This suggested that these residues possessed some 3_{10} content, however only residue Lys_{665} possessed a peak at a distance less than an assumed Hydrogen-bonding threshold of 2.5Å; 1.8-2.5Å is regarded as a typical H-O distance for a hydrogen bond [106]. Analysis of the AMBER-ff03 model revealed peaks in the range 2.146 - 2.657Å for residues Leu_{661}-Trp_{666} and Ser_{668}, with residues Glu_{662}-Asp_{664} lying slightly above the 2.5Å threshold (Fig. 3.10). The position of these bond probability distributions suggested that these residues had some 3_{10} content over the course of the last 5ns. The bond probability distributions for the AMBER-ff99SB model revealed peaks in the range 2.253 - 2.657Å for residues Leu_{660}-Trp_{666} and Ser_{668}, with residues Leu_{661}, Leu_{663} and Asp_{664} lying slightly above the assigned 2.5Å threshold. In fact, only residue Ala_{667} displayed a peak considerably out of range of this threshold (Fig. 3.11). As with the AMBER-ff03 model,

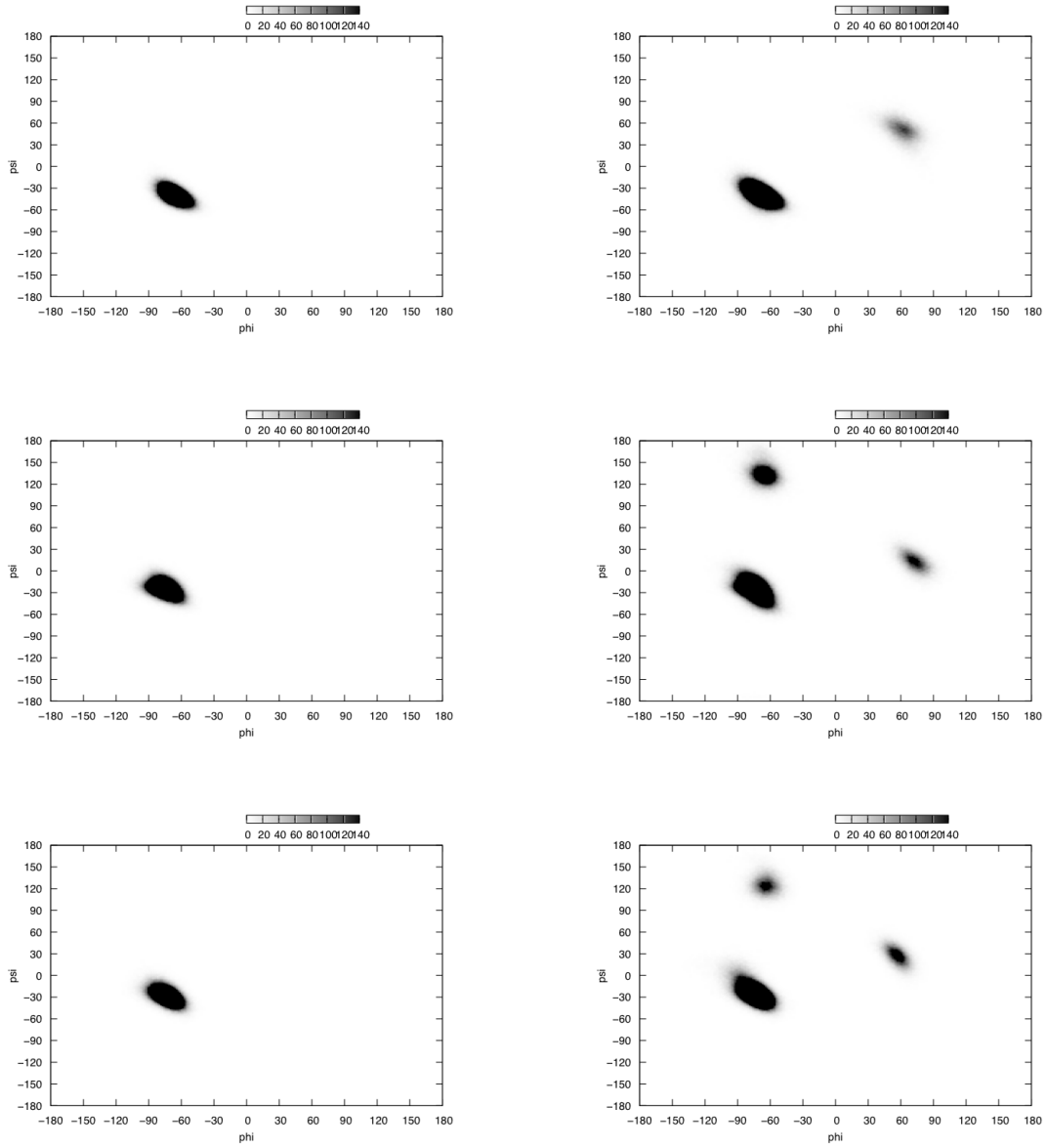


Figure 3.8: Ramachandran plots of the epitope (left) and full sequence (right) over last 5ns of simulation: CHARMM (top), AMBER-ff03 (middle), AMBER-ff99SB (bottom)

the bond probability distributions for i, $i+3$ bonding in the AMBER-ff99SB suggested that residues Leu_{661}-Trp_{666} possessed some 3_{10} content; the difference being that in the case of the AMBER-ff99SB model, residue Leu_{660} also seemed to possess some 3_{10} content.

The bond probability distributions over the epitope for the i, $i+4$ bonds which correspond to the Hydrogen-bonding in an α-helix, revealed that in the CHARMM22 model residues Glu_{662}-Trp_{666} displayed peaks in the range 1.944 - 2.051Å (Fig. 3.9). Additionally, outside of the epi-

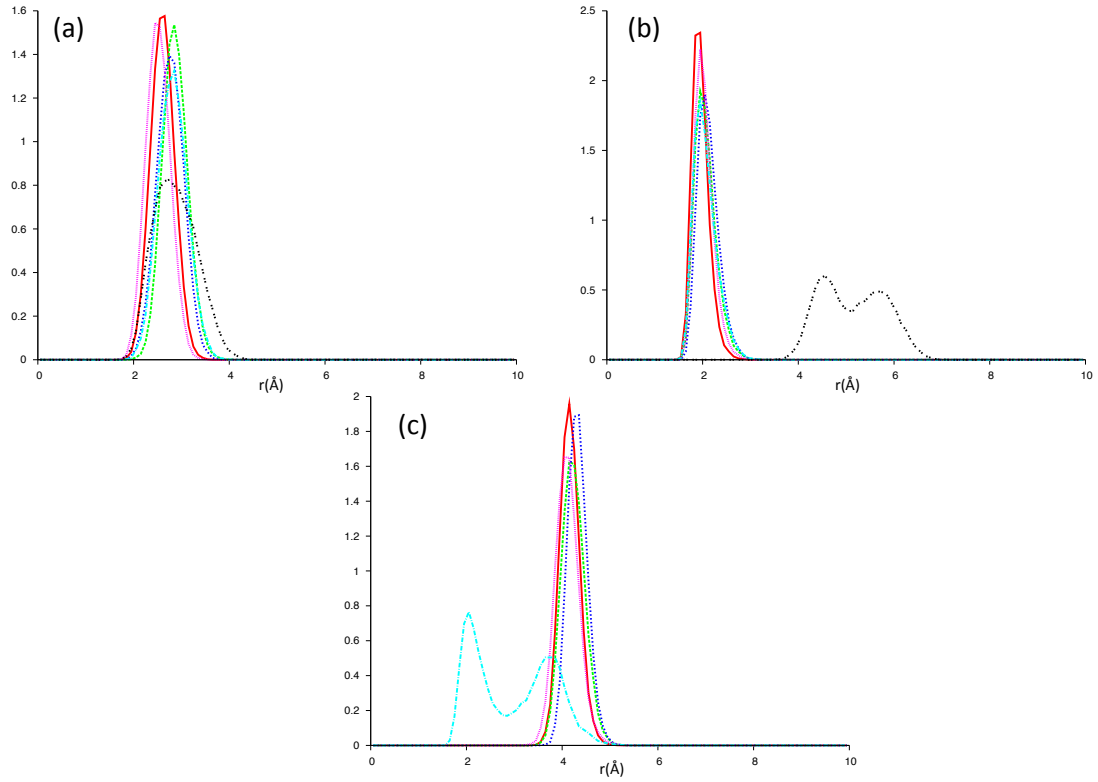


Figure 3.9: Bond distribution over last 5ns of simulation using CHARMM force-field: (a) i,i+3 bonding, (b) i,i+4, (c) i,i+5

tope, residues Leu_{660} and Leu_{661} displayed peaks in the range 2.051 - 2.146Å. Clearly all of these residues possessed peaks in their bond distributions at distances less than the 2.5Å threshold [106], suggesting that these residues possessed a strong affinity for participating in α-helices. The AMBER-ff03 model resulted in bond probability distributions with peaks in the range 2.051-2.265Å for residues Leu_{661}-Asp_{664}. Once again as these peaks were all at distances below the 2.5Å cut-off, this suggested that these residues had a significant α-helical content. The final model AMBER-ff99SB revealed peaks in the bond probability distributions in the range 2.051 - 2.253Å for residues Leu_{660}-Asp_{664}, with residue Lys_{665} giving a peak centred around 2.764Å (Fig. 3.11). These values clearly suggested that residues Leu_{660}-Asp_{664} in particular had a strong α-helical content.

Finally, evaluation of the bond probability distributions for i, $i+5$ bonds corresponding to the Hydrogen-bonding in a π-helix revealed that there was no π-helical content in any of the three simulations (Fig. 3.9, 3.10, 3.11). This was inferred from the lack of successive peaks below 4Å in any of the three systems.

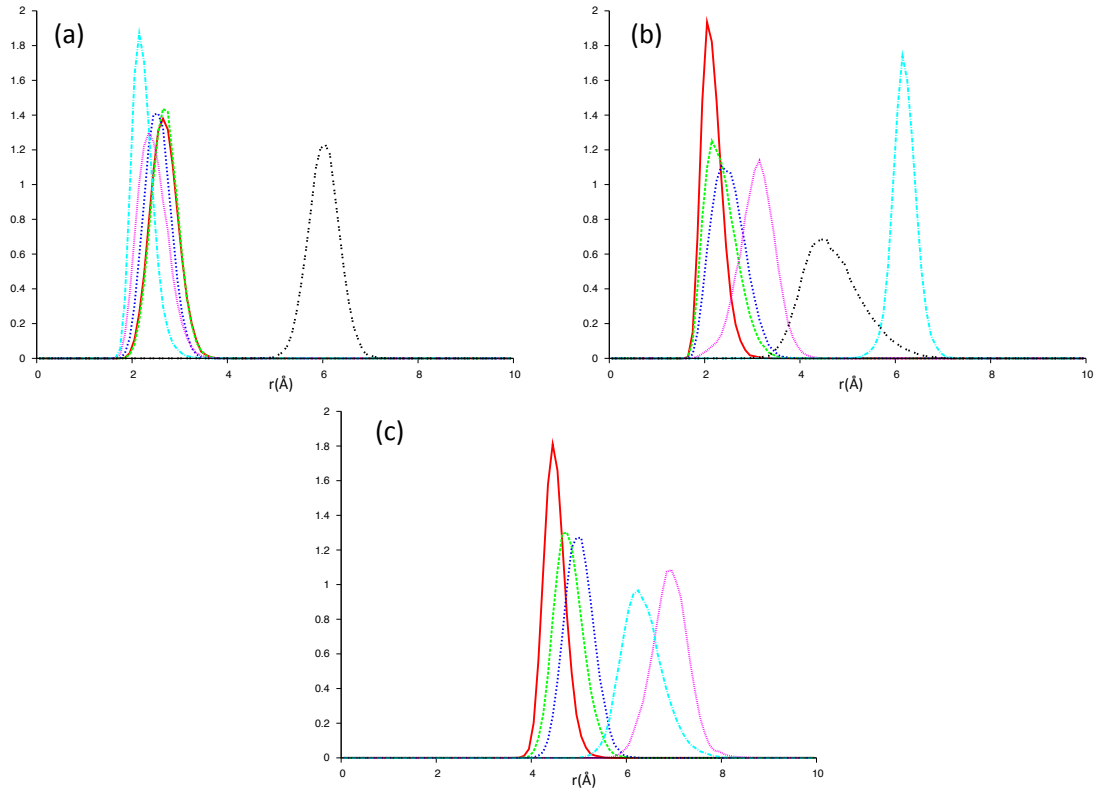


Figure 3.10: Bond distribution over last 5ns of simulation using AMBER-ff03 force-field: (a) i,i+3 bonding, (b) i,i+4, (c) i,i+5

On the basis of the above results one could deduce that in the system using the CHARMM22 force-field, the peptide gp41$_{659-671}$ folded into a helical structure spanning at least residues Leu$_{660}$-Leu$_{669}$, with the Glu$_{659}$ residue transiently extending this helix. This helical structure comprised primarily an α-helix however the rare 3_{10}-helix did appear to contribute to the structure of the peptide to some degree on the basis of bond probability distribution analysis. This conclusion was supported by the torsional angles of the amino acids in question; Leu$_{660}$-Leu$_{669}$ all had torsional angles within the right-handed α-helical region of dihedral space; and the bond probability distribution functions displayed clear i, $i+4$ and potential i, $i+3$ Hydrogen-bonding. In order to highlight the formation and subsequent stabilisation of this helical structure, one can analyse the secondary structure of each residue as a function of time utilising the STRIDE secondary structure assignment algorithms [107] (Fig. 3.12). In the case of the CHARMM22 model, the structure of gp41$_{659-671}$ became dominated by an α-helical structure spanning residues Leu$_{660}$ to Leu$_{669}$ (residue Glu$_{659}$ also participated in this helix for considerable periods).

The secondary structure of gp41$_{659-671}$ obtained under the AMBER-ff03 parameterisation

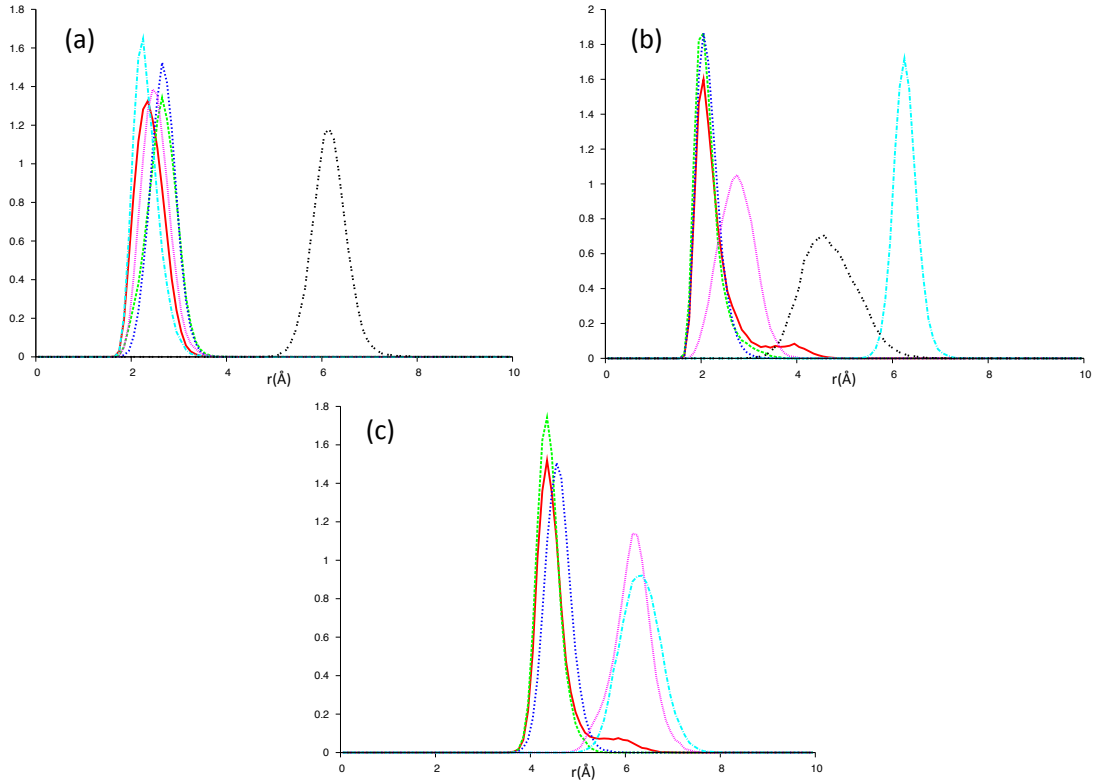


Figure 3.11: Bond distribution over last 5ns of simulation using AMBER-ff99SB force-field: (a) i,i+3 bonding, (b) i,i+4, (c) i,i+5

was that of a helical motif spanning residues Glu_{662}-Ser_{668}. As with the CHARMM22 system, this helix comprised primarily an α-helix with potential 3_{10} content. This was inferred from the torsional angles of these residues all lying within the α-helical region of dihedral space, whereas the torsional angles of residues Leu_{661} and Leu_{669} fell within the region of dihedral space usually attributed to β-strands therefore excluding them from belonging to this helix. The results of the bond probability distribution analysis also supported this hypothesis, showing clear α (and potential 3_{10}) helical Hydrogen-bonding for the residues in question; i, $i+4$ peaks appeared below 2.5Å for residues Glu_{662}-Asp_{664} indicating Hydrogen-bonding up to and including residue Ser_{668}. Analysis of the secondary structure of each residue as a function of time using the STRIDE algorithm supported the interpretation of a helix spanning only residues Glu_{662}-Ser_{668}. This helix was dominated by an α-helix, yet the residues appeared to transiently adopt 3_{10}- and turn-like structure (*cf.* CHARMM model which was entirely α-helical). Out with this sequence, residues Glu_{659}-Leu_{661} have been assigned a stable turn-like structure, and residues Leu_{669}-Asn_{671} have been assigned an aperiodic structure.

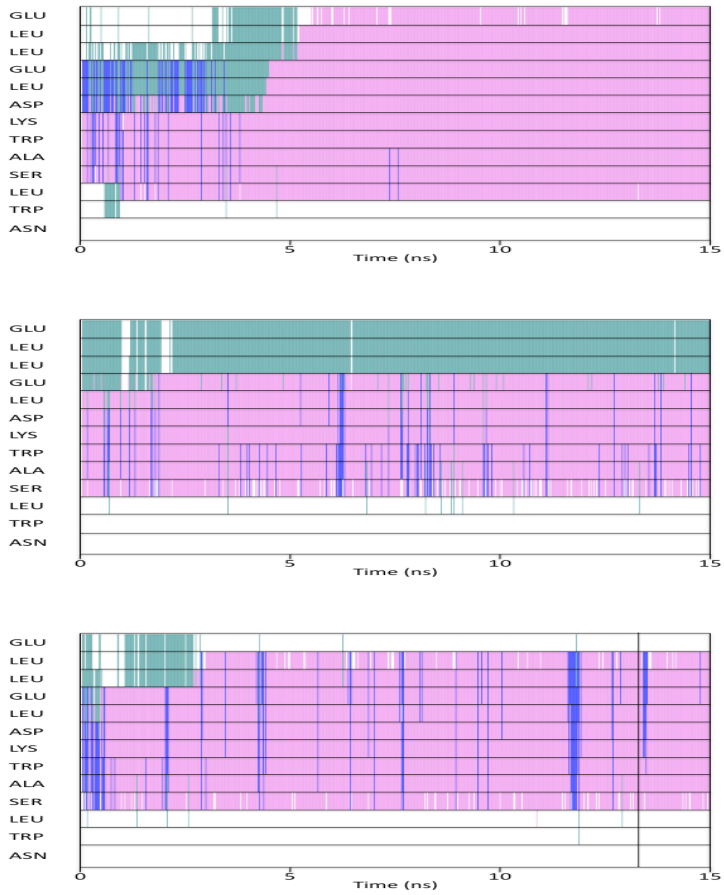


Figure 3.12: Secondary structure of each residue as a function of time: CHARMM (top), AMBER-ff03 (middle), AMBER-ff99SB (bottom). Pink corresponds to α-helix, blue is 3_{10}-helix, green is turn and white corresponds to unordered.

Finally, the secondary structure of gp41$_{659-671}$ from AMBER-ff99SB was also that of a helical motif, spanning residues Leu$_{660}$-Ser$_{668}$. As with the other two models, this helix was dominated by an α-helix with the rare 3_{10}-helix contributing to the secondary structure of the peptide to some degree on the basis of the bond probability distribution analysis for this system. This was supported by the presence of i, $i + 4$ bonds up to a value of i corresponding to residue Asp$_{664}$, indicating a Hydrogen-bonding arrangement consistent with an α-helix up to and including residue Ser$_{668}$. The torsional angles of the residues showed a successive population in the right-handed α-helical region of dihedral space for residues Leu$_{660}$-Ser$_{668}$ also supporting this interpretation. Subsequent analysis of the secondary structure as a function of time using STRIDE also supported this interpretation, having revealed a predominantly α-helical

structure spanning these residues (Fig. 3.12) with the residues out with this sequence adopting an unordered structure. As with the AMBER-ff03 model, the residues comprising this helical structure also appeared to transiently adopt 3_{10}-like structure and for a considerable period appeared to form a stable 3_{10}-helix spanning the entire length of this helical region.

In summary, the simulations indicated that the structure of gp41$_{659-671}$ reported by Barbato *et al.* comprising of no dominant secondary structure, was not a stable motif. In each of the three simulations the epitope of the peptide folded into a helical conformation consisting of a mixture of α (major) and the rare 3_{10}-helices (minor), with the degree of mixing being dependent on the force-field. In each of the three simulations, while the flexibility of the peptide outside of the epitope was force field dependent, the second Trp residue was not assigned an α-helical conformation, confirming that the two Trp residues did not fold into this co-operatively under aqueous conditions. Of the three force-fields used in this study, the CHARMM parameterisation was clearly the most biased towards a helical structure and this finding was certainly consistent with the previous in depth study of classical force-fields [32]. Additionally it would appear that the simulation utilising the AMBER-ff03 force-field gave the best agreement with results obtained experimentally, in that the model showed the greatest content of secondary structure over and above that of a helical motif and suggested an ensemble of conformers at the residue level; the far-UV CD spectra were consistent with a rugged energy landscape indicative of a flexible conformational ensemble. The AMBER-ff03 model also demonstrated that regardless of the absence of i, $i+4$ hydrogen bonding between the two Trp residues, these residues were positioned in order to facilitate simultaneous interaction with a membrane surface (Fig. 3.13). The lack of an α-helical structure spanning these residues may be attributed to this structure being representative of that under aqueous conditions, and as this peptide is membrane-active then this folding may only transpire in close proximity to a membrane surface.

However the results of all three simulations still indicated that these force-fields over emphasised the helical content in pure water when compared to results obtained experimentally. In order to enhance the configurational sampling of these systems, parallel tempering (replica exchange) simulations should be conducted utilising the two AMBER force-fields.

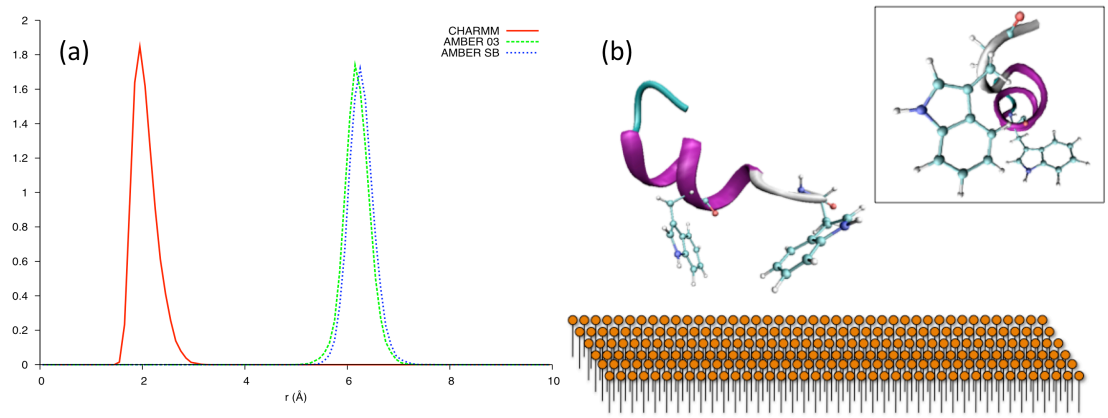


Figure 3.13: (a) Bond distribution between Trp_{666} and Trp_{670} over the last 5ns of each simulation. (b) Proposed interaction of the Trp residues with the membrane surface. The secondary structure predicted by AMBER-ff03 reveals that they are orientated in a way to facilitate binding with a membrane surface. Insert: Side-view reveals that these Trp residues are not perfectly aligned, potentially indicating that further folding may occur on a membrane surface.

3.4.2 Structural variations due to lower dielectric constant

In order to investigate the effect of the microenvironment that the peptide could experience in the proximity of the cell membrane, TFE titrations were performed on the peptide solution. The hydrophilic solvent TFE is widely known to often induce α-helical structure in short peptides [108–110], yet some groups have reported stable β-strands in TFE [111–113]. TFE therefore appears to stabilise the secondary structure for which the amino acid sequence has a propensity [114]. It is also believed to provide a mimic of the microenvironment close to a membrane surface by lowering the solvent dielectric constant [78]. A concentration of 30% TFE (v/v) was chosen as the maximum value as this has been reported to mimic the membrane environment more closely than 100% [115].

In the CD experiments carried out, the ensemble of peptide conformations changed upon addition of TFE in such a way that there was a shift in the conformer population towards higher percentages of helical structures. This was indicated by the formation of a band of positive ellipticity centred around 190 nm, the shift of the negative peak from 199 nm to 205 nm and the appearance of a small shoulder at around 220 nm (Fig. 3.14). The α-helix is known to display large CD bands with negative ellipticity at 222 & 208nm and a CD band with positive ellipticity at 192nm [1, 41–44]. The exact position of the band with negative ellipticity centred around 222nm being heavily influenced by the length of the helix & the dynamical motion of the

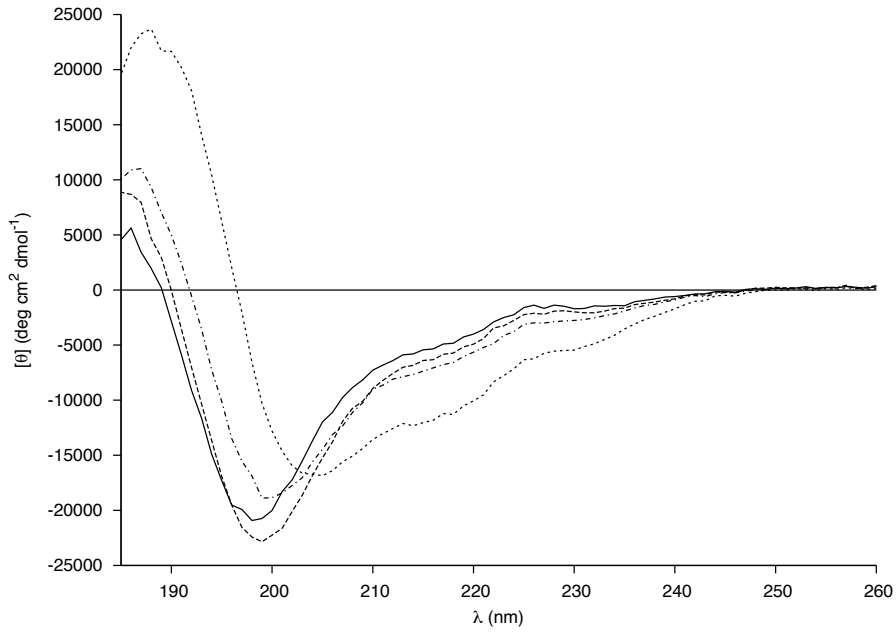


Figure 3.14: Changes in the far-UV CD spectra of 0.04 mM peptide as a function of % of TFE. Line: no TFE; dash: 10% TFE; dash and dots: 20% TFE; dots: 30% TFE. The cuvette pathlength used was 0.2 cm.

protein [116]. Similarly to the α-helix, the 3_{10}-helix displays CD bands with negative ellipticity at 222 & 207nm (the former being a shoulder), with a weak positive band near 195nm. The ratio of $[\theta]_{222}$ / $[\theta]_{207}$ is found to be much weaker than that of the α-helix. An important factor to consider therefore was this ratio over the range of concentrations of TFE, while the % helical content could be followed in a similar fashion (Eq. 3.1). This % was based on the assumption that a value of -33,000 deg.cm^2.dmol^{-1} is taken for 100% helix content [117].

$$\%Helicity = 100 \times \frac{[\theta]_{222}}{33000} \tag{3.1}$$

As evident from these values (Fig. 3.15), there was a clear increase in helical content as the TFE concentration was raised. The ratio of the $[\theta]_{222}$ / $[\theta]_{207}$ values seemed to indicate a structure intermediate between an α- and 3_{10}-helix, however this value did increase with the increasing concentration of TFE suggesting the α-helical structure became more dominant.

Mixed water-alcohol solvents have been reported to model a decrease in dielectric constant effectively, with the effect of various alcohols on a peptide's secondary structure reportedly depending entirely on the average dielectric constant of the solution rather than the individual prop-

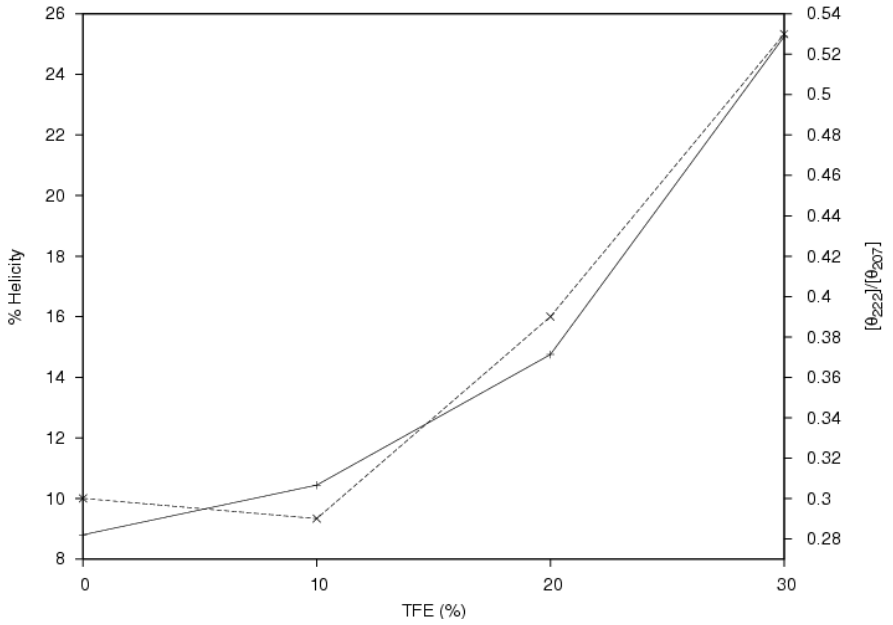


Figure 3.15: Helical content (solid line) and ratio of $[\theta_{222}]/[\theta_{207}]$ (dotted line) of 0.04mM gp41$_{659-671}$ over various concentrations of TFE.

erties of the alcohols [118, 119]. Having exploited the fluorinated alcohol trifluoroethanol (TFE), it was therefore of interest to expose any differences between fluorinated and non-fluorinated alcohols. Additionally, organic co-solvents are known to promote intramolecular hydrogen bonding through the removal of water molecules in the proximity of the solute (encompassing the solute in a hydrophobic 'matrix') and by lowering the dielectric constant. It was therefore also of interest to compare the effect of a solvent with comparable dielectric constant, yet which lacked hydrogen bonding donors. To that end CD spectra of 0.04mM gp41$_{659-671}$ in 20mM Sodium Phosphate were recorded in the presence of either methanol (50% v/v) or acetonitrile (50% v/v); chosen as they had similar dielectric constants to one another and have both been used previously as mimics for membranes [78, 80, 81, 120].

The changes in the CD spectra (Fig. 3.16) were qualitatively similar to those previously observed for TFE, with the ensemble of peptide conformations shifting towards a higher helical content. This was indicated by a red shift in the position of the band of negative ellipticity from 199 to 202 nm, coupled with an increase in the intensity of a negative shoulder centred around 220 nm and the emergence of a positive band around 190 nm. As Trp residues are known to be capable of contributing to the far-UV CD spectra, Fluorescence spectra were recorded in

order to investigate potential changes in the microenvironment of the Trp side chains (Fig. 3.17). N-acetyl tryptophan amide (NATA) was used as a model compound for a completely exposed Trp molecule; such a control method is necessary when comparing spectra of different solvent conditions due to such spectra being extremely sensitive to changes in the environment. Upon addition of each solvent, the changes observed for the peptide were identical to those observed for the NATA control (Fig. 3.17) indicating that the tryptophans were solvent exposed. These results were compatible with an increase in helical population with both Trp residues favourably orientated to interact with a membrane (Fig. 3.13).

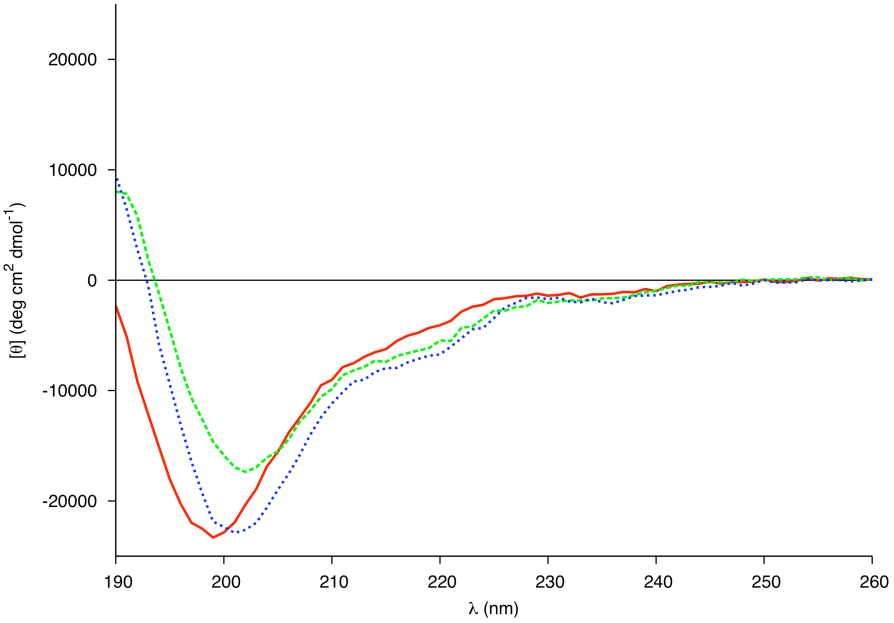


Figure 3.16: Far-UV CD spectra of 0.04mM gp41$_{659-671}$ in 20mM sodium phosphate buffer pH 6.83 (Solid line), in 20mM sodium phosphate buffer and 50% Methanol (Dashed line) and in 20mM sodium phosphate buffer and 50% Acetonitrile (Dotted line).

The mixture of 50% methanol (v/v) resulted in a helical composition of 13%, which was half of the value obtained for 30% TFE (Fig. 3.15). This may have been due in part to the differences in the dielectric constant values for these two solvents (with the net dielectric constant of the mixed solvent being lower in the case of TFE) and/or to the differences in stabilisation of secondary structure elements by fluorinated solvents [121]. A helical content of 15.9% was obtained for the mixture of 50% acetonitrile (v/v). As acetonitrile and methanol have a similar dielectric constant, this small increase in helical fraction may have been evidence of the solvents lack of hydrogen bond donors further promoting the formation of intramolecular hydrogen bonds.

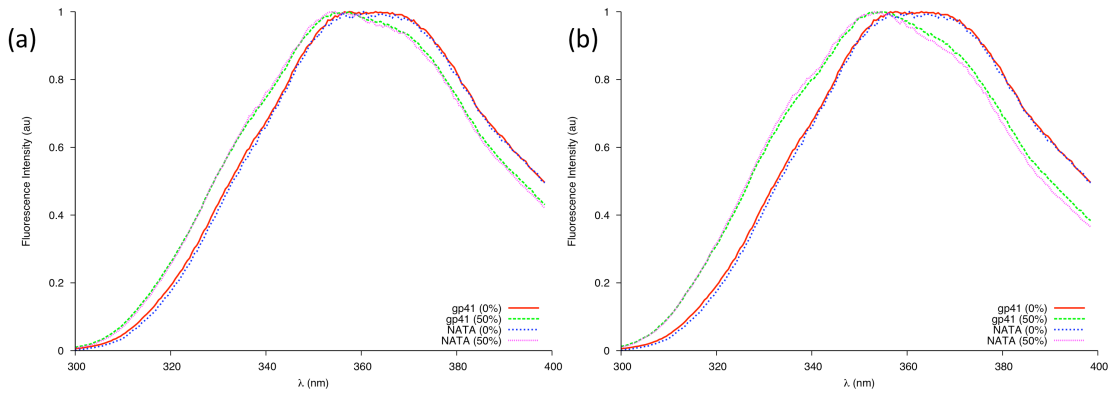


Figure 3.17: Changes to the intrinsic Trp fluorescence of the gp41 peptide and 10μM NATA control due to the presence of (a) 50% Methanol and (b) 50% Acetonitrile.

The solvents used in this mixed-solvent study were all less polar and had a lower dielectric constant than water. As the dielectric constant has a linear relationship to the number of hydrogen bonds in solution, both of these factors reduce the total number of hydrogen bonds between solvent and peptide. This appeared to promote the formation of intramolecular hydrogen bonding within the peptide, promoting the helical conformation in each of the solvents studied. As these mixed solvents can mimic the extended regions of a membrane [81], it is therefore natural to conclude that the secondary structure of the membrane proximal peptide $gp41_{659-671}$ is more ordered in its native state near a membrane surface than under simple aqueous conditions. These results pertaining to the lowering of dielectric constant, appeared to be in greater agreement with the results of the MD simulations than those obtained under aqueous conditions. This may be evidence of the force fields over-expressing the helical content in pure water potentially due to an inability to correctly handle the intermolecular and intramolecular competition for hydrogen bonding. The AMBER-ff03 model may in reality represent the secondary structure of this sequence in an environment with a lower dielectric constant than that of water, with the helical conformation spanning residues Glu_{662}-Ser_{668} being representative of the increase in helical composition observed in the CD spectra (Fig. 3.16).

3.4.3 Mimicry via micelles

Here the cumulative partial denaturation effects of a membrane surface were replicated; that of the negative charges of a membrane [77] and the decrease in dielectric constant near the membrane surface [78–80]. Far-UV CD spectra of 0.04mM $gp41_{659-671}$ in 20mM sodium phosphate

buffer were recorded in the presence of varying concentrations of Sodium Dodecyl Sulphate (SDS). Micellar SDS provides an anionic membrane-mimicking environment [122] and has been used extensively to study membrane peptides [123–127]. The simplicity of an SDS micelle (as opposed to liposomes or vesicles) makes it an ideal starting point to study the secondary structure of the membrane-proximal $gp41_{659-671}$ and to elucidate the interactions between the peptide and membrane interfaces [98].

In the CD experiments carried out, the ensemble of the peptide conformation changed upon addition of SDS with a shift in the population towards higher percentages of helical structures (Fig. 3.18). This was indicated by a red shift in the band of negative ellipticity, the formation of a positive band around 190 nm and the emergence of a shoulder centred around 227-228 nm. This latter band appeared at higher wavelengths when compared to the classical 220 nm position of the α-helix, suggesting a major contribution to the CD spectra from the Trp residues (known to give rise to bands in this region [128]). As this band has so far only appeared in the presence of SDS, this may be evidence of the two Trp residues folding into the proposed helical structure in order to directly interact with the micelle surface. Fluorescence spectra of $gp41_{659-671}$ in the presence of micelles revealed that the microenvironment of the Trp residues was indeed altered (Fig. 3.19a), with the shift towards lower wavelengths being indicative of the microenvironment of the Trp residues becoming less polar [71]. As this difference between the NATA control and the peptide was only found in the presence of micellar SDS (Fig. 3.19b), it implied membrane-mediated folding of the peptide upon the interaction of the tryptophan residues with the micelle.

The first step of the interaction between Trp-rich antimicrobial peptides and membranes is an electrostatic association between these positively charged peptides and the negatively charged micelles [129]. As the sequence $gp41_{659-671}$ carried an overall charge of -2 (localised towards the N-terminus of the peptide) at neutral pH, it was important to investigate the electrostatic interactions between this peptide and the negatively charged head groups of SDS *via* variations in the pH of the sample. Far-UV CD spectra of $gp41_{659-671}$ in 20mM sodium phosphate and 7mM SDS under neutral, acidic and basic conditions were subsequently recorded (Fig. 3.20).

Under acidic conditions, SDS-induced α-helical folding appeared to be further enhanced. This was evident through the increased intensity in the band of positive ellipticity and the appearance of an additional negative shoulder near 220nm; no such increase was detected in the CD

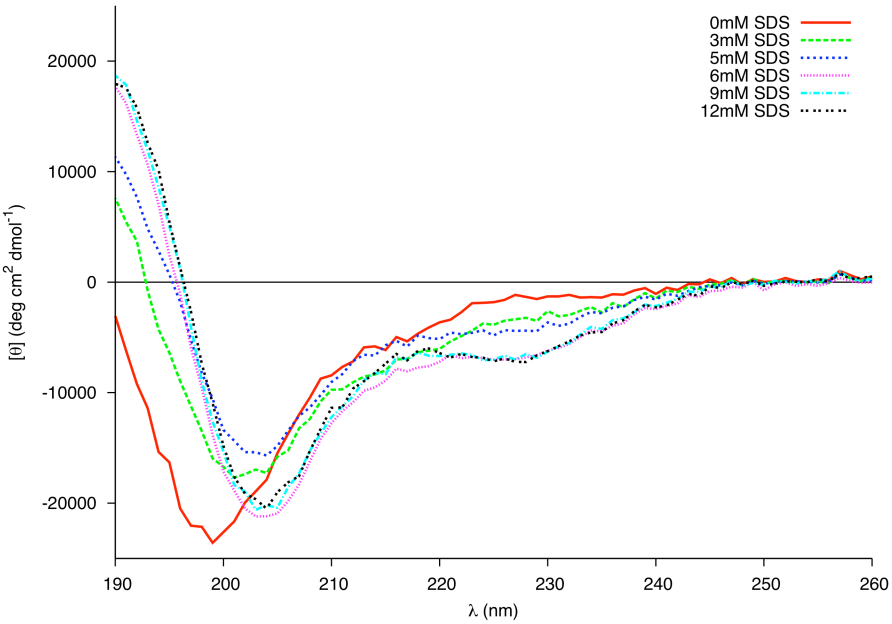


Figure 3.18: Changes in the far-UV CD spectra of 0.04 mM peptide in 20mM Phosphate buffer due to increasing concentrations of SDS at 25°C. Above the cmc all changes are completed.

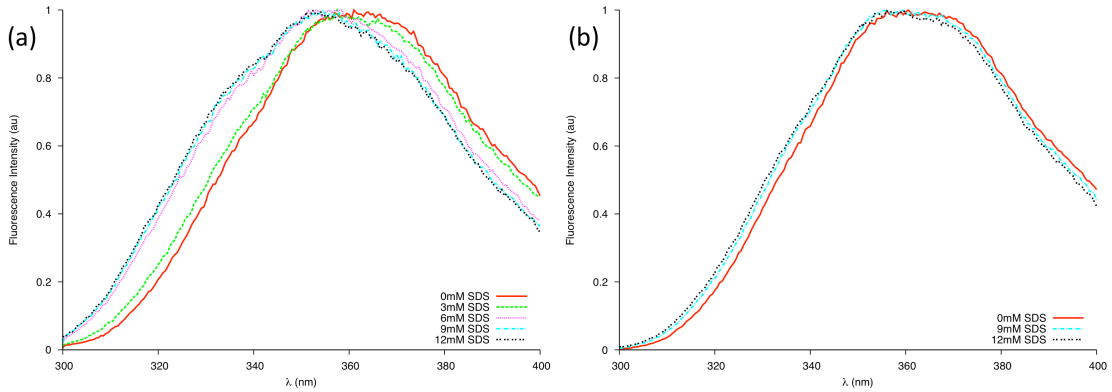


Figure 3.19: (a) Intrinsic Trp fluorescence of the gp41 peptide in the presence of different concentrations of SDS below (3mM) and above (6, 9, 12 mM) the cmc. (b) Intrinsic Trp fluorescence of 10μM NATA in the presence of different concentrations of SDS.

spectra of 0.04mM gp41$_{659-671}$ in 20mM sodium phosphate in the absence of SDS (Fig. 3.21). However under basic conditions this SDS-induced folding was weakened, as the major negative band was blue shifted and accompanied by a decrease in the intensity of the positive band centred around 190nm.

This structural sensitivity due to pH was only manifested in the presence of SDS micelles; the presence of micelles for all pH values was confirmed using dynamic light scattering (Fig. 3.2b). It was therefore surmised that the molecular mechanism responsible was due to the electrostatic

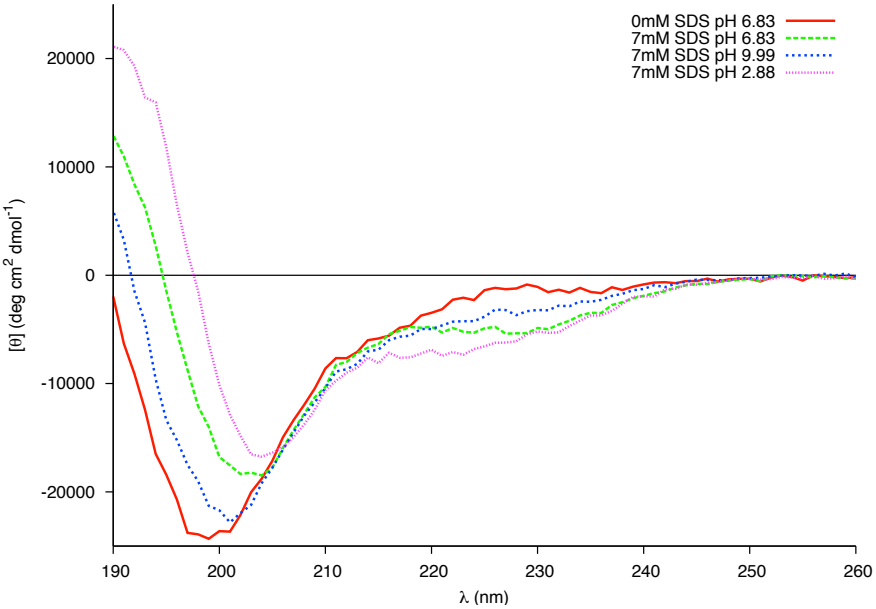


Figure 3.20: Changes in the far-UV CD spectra of 0.04 mM peptide in 20mM Phosphate buffer and 7 mM SDS at different pH.

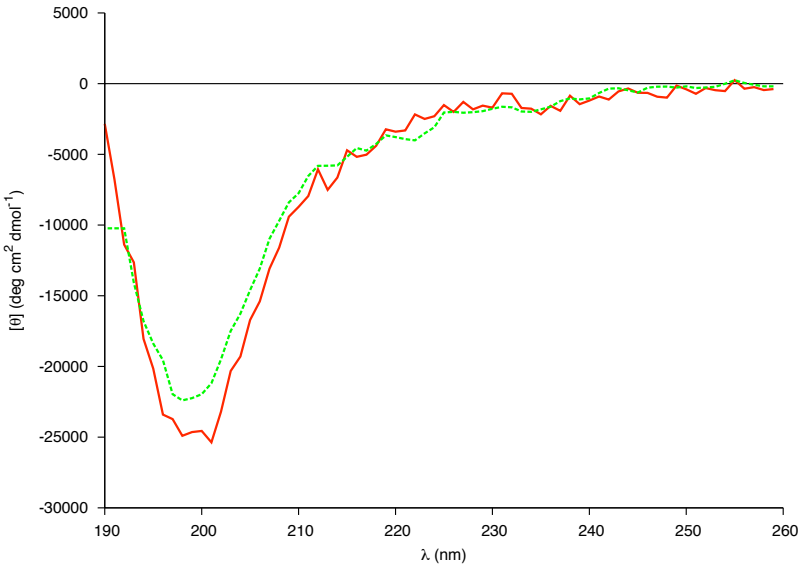


Figure 3.21: Changes in the far-UV CD spectra of 0.04 mM peptide in 20mM Phosphate buffer at pH 6.83 (solid line) and pH 2.14 (dashed line).

interaction between the negatively charged head groups of SDS and the charged residues of $gp41_{659-671}$. As mentioned above, under neutral conditions it appeared that the Trp residues

(positioned at i and $i+4$) folded into an α-helix in order to interact with the membrane surface. A strong electrostatic interaction between the Glu_{662} and Lys_{665} residues (potentially resulting in a competing i, $i+3$ electrostatic intra-helical interaction) would also be present; such an interaction was supported by all three of the MD simulations which revealed potential 3_{10} hydrogen bonding between these residues (Fig. 3.22).

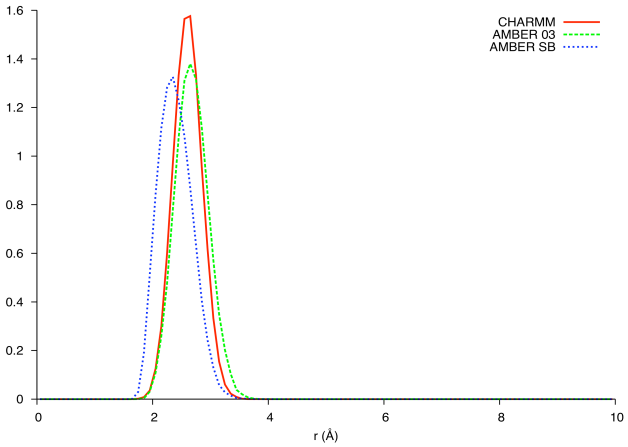


Figure 3.22: Bond distribution between Glu_{662} and Lys_{665} over the last 5ns of each simulation.

As the two Trp residues bind to the surface of the micelle, the positively charged Lys residue would be attracted to the negatively charged surface and may potentially act as an 'anchor' for this helix. There would subsequently be a strong electrostatic repulsion between the anionic head groups of SDS and the Glu_{662} residue, destabilising this 3_{10}-helix. Coupled with the strong electrostatic repulsion between the anionic SDS head groups and Asp_{664}, this would result in the N-terminal part of $gp41_{659-671}$ adopting a more frayed-like structure directed away from the micelle surface. Under acidic conditions, the carboxylic acid groups of the Glu and Asp residues would be protonated, therefore removing the electrostatic repulsion between the negatively charged groups of the peptide and micelle. This protonation subsequently results in the co-operative folding of ELLELD into a helical structure with the sequence spanning the two Trp residues. Finally under extreme basic conditions, deprotonation of the Lys residue also removes the i, $i+3$ electrostatic interaction and the electrostatic repulsion between the anionic species increasing and affecting the ability of the two Trp residues to successfully bind to the micelle. This is certainly consistent with the results obtained for the full MPR sequence [94], and suggests that the pH plays an important role in the conformational stability of $gp41_{659-671}$ at the

membrane interface.

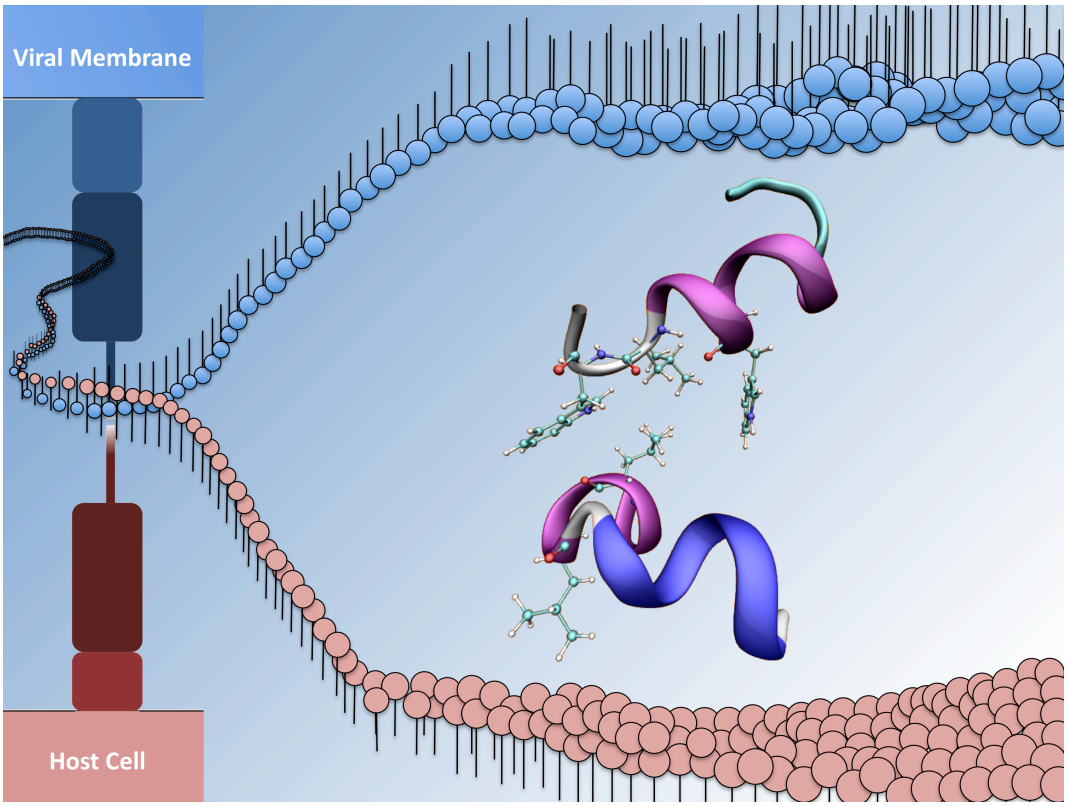
Interestingly, an independent study investigating pH dependence on the secondary structure of a peptide (both under aqueous and micellar SDS environments) found that while the aqueous structure was highly dependent on the pH of the system, the secondary structure promoted by the micelle did not vary greatly under the pH conditions tested [130]. Such a result was claimed to be expected due to the interaction with SDS potentially imposing constraints on any structural effects caused by variations in the pH. By contrast, we found appreciable pH tuneability suggesting that $gp41_{659-671}$ is structurally flexible even in close proximity to a membrane surface.

3.5 Conclusions

The structural properties of the HIV peptide $gp41_{659-671}$ have been studied using mixed solvents, anionic micelles and MD simulations in order to mimic the distinct local environments relevant to viral fusion. The results revealed $gp41_{659-671}$ presented an independent capability to interact with a membrane surface, with different segmental secondary structures being adopted during the fusion process. The bias towards helical folding could be tuned by pH variation to the extent that the membrane proximal peptide showed no defined fold type at basic pH and that it was helical near physiological and lower pH environments. This was in contrast to results obtained in the absence of membrane mimicking SDS, which revealed no such control of helical folding by pH under aqueous conditions. The pronounced sensitivity to standard empirical potentials was also explored and it was concluded that AMBER-ff03 provided a reasonably accurate description of the solution state structures and provided a marked improvement over the previously employed CHARMM force field, which substantially overestimated the helical population. On the basis of this force field study, it would be interesting to determine if the AMBER-ff03 force field transfers to the membrane environment and provides useful input to aide the interpretation of the experimental results.

4

Membrane binding: Structure variations of $gp41_{532-544:659-671}$ in the hemifusion phase.



4.1 Rationale

In the previous chapter, the flexibility of the membrane proximal region (MPR) of the HIV-1 gp41 envelope was investigated by focusing on a very short MPR fragment, $gp41_{659-671}$. The flexibility of this region is believed to be relevant to its biological function, therefore it is important to consider what role gp41 fills in the infection process of HIV.

A critical step in this infection process is the gp41-mediated fusion between the viral envelope and the cellular membrane, which is driven by a sequence of conformational changes in the gp41 ectodomain (i.e. the domain that extends into the extracellular space and initiates contact with the surface of the host cell). Current evidence suggests that the protein initiates host membrane binding through a pre-hairpin intermediate, serving as a bridge between the two membranes. In this pre-fusion phase, three gp41 monomers are held together in parallel by a trimeric coiled-coil domain formed by the N-terminal heptad repeat 1 (HR1) regions; a heptad repeat consists of a repeating pattern of seven amino acids. It is this phase which is attractive for therapeutic intervention as the relatively long conformational exposure of the structure renders it rate-limiting [131]. As such, primary anti-HIV strategies constitute peptides and neutralising antibodies targeting HR1 in order to inhibit the helical transition of the C-terminal heptad repeat 2 (HR2) region and the refolding of the pre-hairpin into a core structure designated the six-helical bundle (6HB). This transition catalyses the fusion process by forcing the viral and cell membranes into close proximity with one another, resulting in hemifusion or stalk intermediates. The formation of these mixed lipidic structures suggests that two membrane proximal regions, the C-terminal membrane proximal external region (MPER) and the N-terminal fusion peptide proximal region (FPPR), may have a significant and potential regulatory role in the fusion process. These two regions link the 6HB to the cell and viral membrane-inserted domains, termed the fusion peptide (FP) and transmembrane region (TMR) respectively (Fig. 4.1), and are believed to be involved in the lipid-mixing phase.

It is reasonable to hypothesise that the two domains that are aligned within the 6HB should contain peptide sequences which exhibit mutual interactions dependent upon their membrane environments [132, 133]. Before fusion, these stretches are in close proximity within the envelope spike (Env) and share the viral membrane environment; within the envelope spike, the

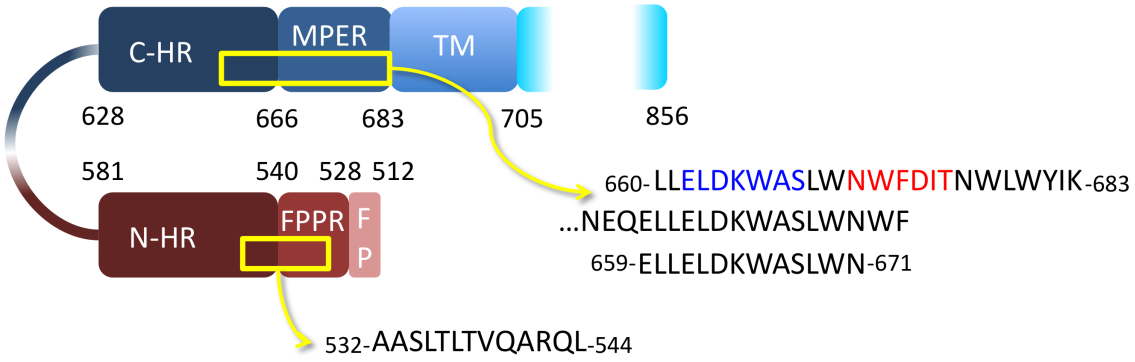


Figure 4.1: Schematic representation of gp41 regions in the 6HB, revealing the peptide sequences of interest.

gp41 ectodomain is wrapped by the gp120 trimer. These stretches then separate during the pre-fusion phase, before coming together again in a different environment (including the host cell membrane) upon the formation of the 6HB [132, 133].

The conserved primary structure of MPER sequences has prompted suggestions that the MPER exhibits autonomous interactions with membrane interfaces, with recent investigation demonstrating the membrane insertion of the MPER within the gp41 trimer [3]. The crystal structure of gp41 includes both proximal regions, and implies that these regions may arrange into an ordered (and possibly aligned) assembly beyond the core thereby contributing to the fusion reaction. Furthermore, the hydrophobic residues of the MPER have been observed to be binding determinants in the interactions with the FPPR [133], which would appear to be supported by the investigation into $gp41_{659-671}$ in the previous chapter; the results of which suggested that the Trp_{666} and Trp_{670} residues were important in MPER-membrane interactions, and the stretch was capable of self-regulated membrane binding in response to pH and potentially (due to the involved electrostatics) to the membrane composition itself. It is therefore of interest to probe whether this membrane activity could be co-regulated with a matching FPPR stretch.

A series of studies have been aimed at the elucidation of the structure-activity relationships of the fusion inhibitor DP178 (or T20) peptide, which spans the $gp41_{643-678}$ stretch comprising the HR2 and a portion of the MPER [134, 135]. Interestingly, it would appear that it is the MPER portion of this peptide which determines not only the membrane interactions of T20, but also its inhibitory potency [94, 135–138]. T20 is known to inhibit membrane fusion at a post-lipid mixing phase, yet lacks the coiled-coil binding residues of HR1 which are believed to be

necessary for preventing HIV-1 entry. Furthermore, T20's inhibitory activity directly correlates with its ability to bind at membrane surfaces [139] and with its enhanced interactions with the 6HB incorporating the $gp41_{530-544}$ FPPR region [136], which is consistent with recent studies suggesting T20 strongly binds to the $gp41_{532-546}$ region [137]. In total, this data would imply there is a specific membrane-bound inhibitory stretch for T20 within the FPPR, and on the basis of the presented stretches, it would appear to be the $gp41_{532-544}$ region.

4.2 MPER-FPPR stretch pair selection

Having previously investigated the conformational preferences of the sequence $gp41_{659-671}$ in *Chapter 3*, it stands to reason that this region would define the choice of stretch pair. The $gp41_{659-671}$ sequence itself, which spans the N-terminal portion of the MPER and part of the HR2, has been identified as the epitope for the mAB 2F5 believed to block MPER-FPPR interactions in the hemifusion phase [133, 134, 140, 141]. Furthermore, this sequence has been comprehensively studied under a variety of conditions including solution, micellar and phospholipid environments [2, 86, 91, 94, 97, 138, 141]. Within the crystal structure of gp41, $gp41_{659-671}$ was found to be aligned with $gp41_{532-544}$ (Fig. 4.2), which was regarded as the prime choice for the FPPR stretch. Both the $gp41_{659-671}$ and $gp41_{532-544}$ stretches are terminal extensions of their corresponding HR regions, and it is theorised that the degree to which they fold into a helical structure is regulated by the 6HB.

As with $gp41_{659-671}$ and the FP ($gp41_{512-528}$), the FPPR stretch was expected to be conformationally polymorphic and display environmental sensitivity [142, 143]. Studies utilising the N- and C-terminal fragments of the FP-FPPR sequence ($gp41_{512-541}$) have suggested that the FP and FPPR interact independently from one another with membranes at various steps during the fusion process [134, 144]. Therefore it was of considerable interest to probe the folding of the C-terminal portion of the FP-FPPR (i.e. $gp41_{532-544}$), both individually and in combination with the "complementary" $gp41_{659-671}$. Such an investigation was hoped to reveal the extent to which the interplay between the two sequences could be modulated by changes in membrane model environments.

The $gp41_{532-544}$ peptide is a highly conserved motif across all isolates of HIV-1,HIV-2 and

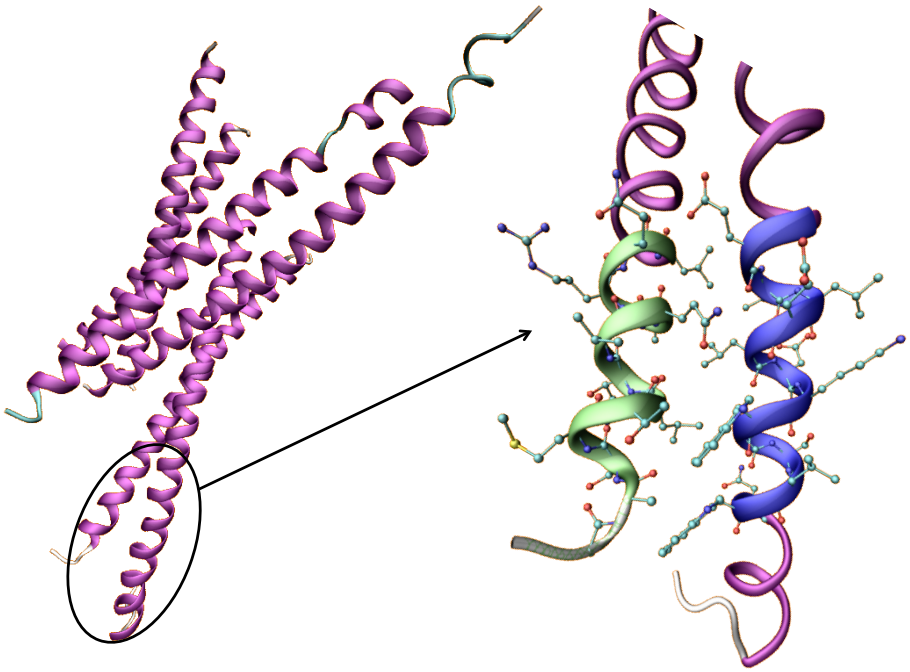


Figure 4.2: Crystal structure of 6HB [3], revealing close proximity of $gp41_{532-544}$ (Green) and $gp41_{659-671}$ (Blue).

SIV and carries a general sequence *H*TLTV, where *H* ($gp41_{535}$) is hydrophobic and predominantly Leu or Ile [133, 145], and in some instances Met (e.g. B-HXB2R) [3]. This motif is regarded as being prone to β-sheet formation, suggesting conformational bias upon membrane binding.

4.3 Model environment

In order to examine the interactions between the proximal regions during the fusion process, it was therefore important to study such systems within a membrane environment. Previous lipid studies of the gp41-mediated HIV fusion involved specific gp41 regions and focused on investigating conformation-stabilising interactions [146, 147]. Additional studies have concerned the impact of lipids on mAbs targeting gp41: interactions involving the mAb 4E10 were found to extend beyond those with the MPER epitopes and relied upon binding with the viral membrane [148–150], while the binding of the mAb 2F5 was reported to be stabilised by an extra docking site in the membrane [150–152]. As this investigation contained herein focused on the interaction interplay upon membrane binding, the reciprocal response of the two stretches were

investigated under different membrane environments. As the stretches are located in proximity to the viral ($gp41_{659-671}$) and host cell ($gp41_{532-544}$) membranes during the pre-hairpin phase, zwitterionic and anionic phospholipid vesicles (mimicking mammalian and viral membranes respectively) were chosen to emulate the lipid environments during the pre-fusion and hemifusion phases.

4.4 Materials & Methods

4.4.1 Synthesis of peptides and membranes

Peptide synthesis was performed by the *National Physical Laboratory* (NPL) on a Liberty system (CEM Corporation). The peptides were assembled on a rink amide MBHA resin (100 μM scale) using standard solid-phase *N*-(9-fluorenyl)methoxycarbonyl (Fmoc) -based protocols with HCTU/DIPEA as coupling reagents. Synthesised peptides were purified by RP-HPLC following post-synthetic deprotection and work-up. The final constructs were identified by MALDI-ToF mass-spectrometry with α-cyano-4-hydroxycinnamic acid as the matrix.

Dilauryl phosphatidylcholine (DLPC) and its mixtures with dilauryl phosphatidylglycerol (DLPG) (Avanti Polar Lipids) were used by the *National Physical Laboratory* (NPL) to prepare zwitterionic and anionic membranes respectively using previously published protocols [153, 154]. Unilamellar vesicles of DLPC and DLPC-DLPG (3:1 molar ratio) were prepared by dissolving dry lipids in chloroform/methanol (2:1 v/v) followed by evaporating the solvents and hydrating the residue to 10 mg ml^{-1} total lipid concentration in 10 mM phosphate buffer, pH 7.4. The suspension was then extensively vortexed, sonicated (30°C), filtered (up to twenty times, polycarbonate filter, 0.1 μm) to give a clear solution, and analysed (100 nm) by photon correlation spectroscopy.

4.4.2 UV-visible spectroscopy

The concentrations of the peptide solutions were determined spectrophotometrically using the calculated extinction coefficients at 214 nm for $gp41_{532-544}$ (ϵ_{214} = 11852 M^{-1} cm^{-1}) and 280 nm for $gp41_{659-671}$ (ϵ_{280} = 14000 M^{-1} cm^{-1}) - calculated from established techniques [155] and used previously [91, 156].

4.4.3 Circular Dichroism (CD) measurements

Far-UV CD spectra were recorded on a Jasco J-810 spectropolarimeter (Japan Spectroscopic Co., Tokyo) fitted with a Peltier unit for temperature control. The spectra were the average of three accumulations in step mode (data pitch of 1 nm, response time of 8s, bandwidth of 1 nm) acquired in rectangular cuvettes of 0.1 cm path length (unless otherwise stated) at 25°C. For every sample, an appropriate solvent spectrum was recorded under identical conditions and subtracted from the corresponding sample spectrum. The data collected was expressed in molar ellipticity (deg.cm^2.dmol^{-1}) and spectra recorded in the presence of synthetic membranes were for the lipid/peptide ratio of 40:1 (1.6mM lipid).

CD spectra with an extended range of wavelengths were acquired at Diamond Light Source (Beam-line 23) using a synchrotron radiation CD spectropolarimeter. CD spectra for individual peptides and their equimolar mixtures were acquired at 0.2 mM (total peptide concentration) using a quartz cuvette with 0.02 cm path length at the stated lipid/peptide ratios and solvent conditions. Experimental mixtures were compared with spectra derived from linear combination of the individual spectra (Eq. 4.1) assuming no perturbation to structure upon mixing.

$$\theta_{x(TheoreticalMix:A+B)} = \frac{\theta_{x(PeptideA)} + \theta_{x(PeptideB)}}{2} \tag{4.1}$$

All measurements performed at Diamond were taken in *mdeg* and for individual peptides, converted to mean residue ellipticity, [θ]. Spectra for gp41$_{532-544:659-671}$ were not converted in order to provide a clear comparison with the linearly combined individual spectra.

4.4.4 Fluorescence Spectroscopy

Tryptophan fluorescence was measured with a LS55 (Perkin Elmer Instruments, Massachusetts, USA) at 25°C using a 1cm quartz cuvette (Hellma GmbH, Germany) and 0.04mM protein solution. To selectively excite the tryptophan an excitation wavelength of 295 nm (2.5nm excitation slit) was used. The spectra were acquired from 300 to 400 nm (emission slit 6nm) and subtracted with the corresponding buffer blank.

4.4.5 Computational Methods

Two unit cells were set up, each comprising of one $gp41_{532-544}$ molecule with zwitterionic termini, 1853 water molecules and 1 chloride counter ion in order to ensure charge neutrality. The starting configurations of $gp41_{532-544}$ were based on the crystal structure of the 6HB (Fig. 4.2) and on the unordered configuration reported by Barbato *et al.* [2] and employed previously for $gp41_{659-671}$ [156]. The latter was achieved through mutation of the $gp41_{659-671}$ sequence through inbuilt features of the Visual Molecular Dynamics (VMD) software [157]. A further unit cell was set up, which comprised one $gp41_{532-544}$ molecule and one $gp41_{659-671}$ molecule (both with zwitterionic termini and based on the unordered configuration) in an anti-parallel arrangement, 1853 water molecules and 1 sodium counter ion in order to ensure charge neutrality.

In each simulation, the molecules were arrayed on a cubic lattice with random molecular orientations in a cubic box of length 38.2Å. The rigid, non-polarisable, three-site TIP3P empirical force field was used to describe the water molecules, the AMBER-ff03 force field was employed to describe the peptides, while the PINY simulation package was employed to perform the simulations [99]. The system was equilibrated at 300K in the canonical ensemble for 150ps with a time-step of 0.5fs to anneal out unphysical contacts. The system was then run for 150ps with a time step of 0.5fs in the isothermal-isobaric ensemble to allow for spatial relaxation. The system was then run for a further 150ps in the canonical ensemble with a time-step of 1fs being employed, before a production run of 15ns using a time-step of 2.5fs from which all data was collected. Periodic boundary conditions were employed and long-range interactions were evaluated via Ewald summation.

4.5 Results

4.5.1 $GP41_{532-544}$

Before moving onto the gp41 peptides in the presence of membrane mimetics, it was imperative to study the structural properties of $gp41_{532-544}$. To that end, structural investigation of the FPPR $gp41_{532-544}$ was conducted under identical conditions to those used for the MPER counterpart $gp41_{659-671}$, chosen in order to mimic the distinct physical and chemical environments that

this sequence may be exposed to during the viral fusion process; under aqueous and mixed solvent conditions mimicking the diffuse layers of varying dielectric constant, and with surfactant solutions mimicking the interfacial contact regime.

Under aqueous conditions, $gp41_{532-544}$ did not fold into a dominant secondary structure with the shape of the recorded far-UV CD spectra indicating an unordered confirmation (Fig. 4.3) and, as with the far-UV CD spectra of $gp41_{659-671}$, was consistent with a flexible conformational ensemble. As such an ensemble could be characterised by an equilibrium between different conformers at a chain level or individual amino acid residues along the sequence adopting, transiently, ϕ and ψ angles typical of different secondary structural elements, it was important to determine if finer structural detail could be explored utilising molecular dynamics simulations.

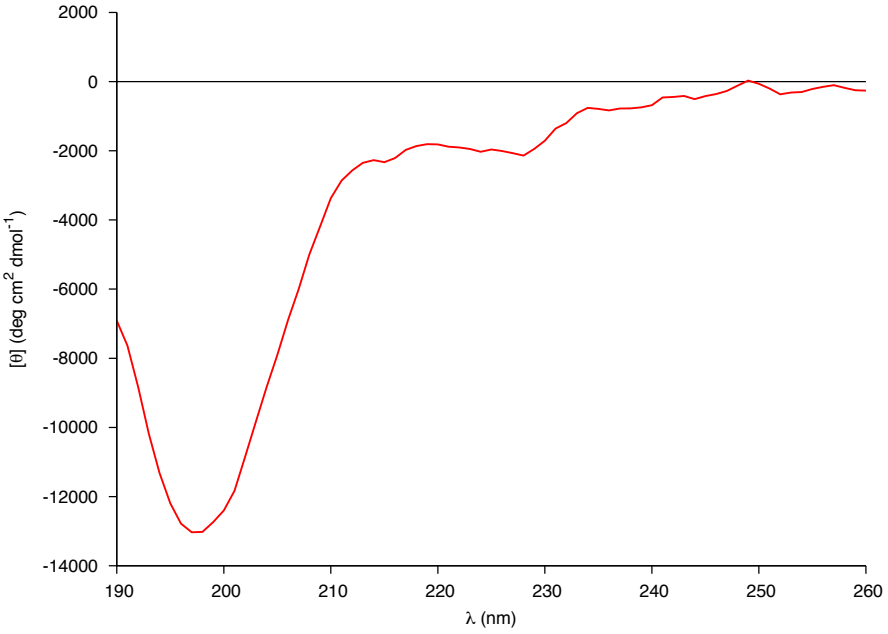


Figure 4.3: Far-UV CD spectra of 0.04mM $gp41_{532-544}$ in 10mM phosphate buffer.

Visualisation of the simulation of the initial helical $gp41_{532-544}$, revealed that the structure had not varied significantly from its initial configuration (Fig. 4.4a), and was defined by a slightly curved helix. Analysis of the secondary structure of each residue revealed that the simulation predicted a stable α-helical motif spanning the residues Ala_{532}-Arg_{542}, and transient 3_{10}-helical structure over the Ala_{541}-Gln_{543} residues (Fig. 4.5). This was supported by the backbone torsional angles for the peptide over the last 5ns of the simulation, which presented a compact peak within the right-handed α-helical region of dihedral space (Fig. 4.4b).

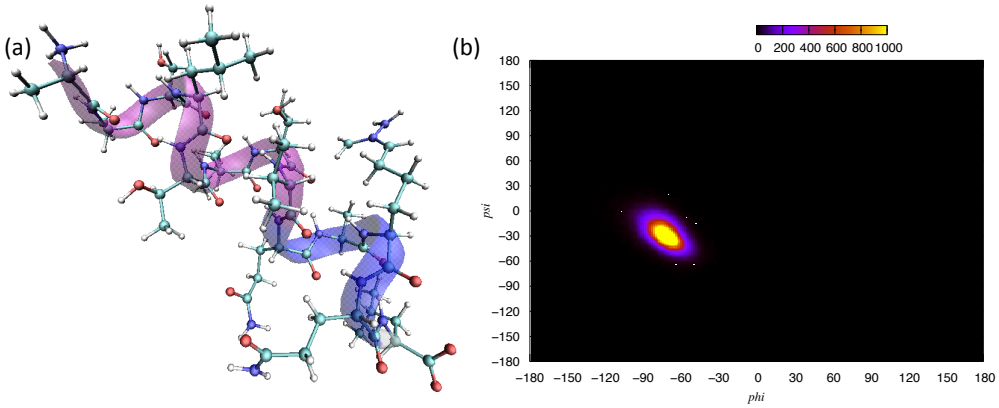


Figure 4.4: (a) Structure after 15ns of initially 'helical' simulation of $gp41_{532-544}$, revealing continuance of helical motif. (b) Ramachandran plot of the simulation over the last 5ns.

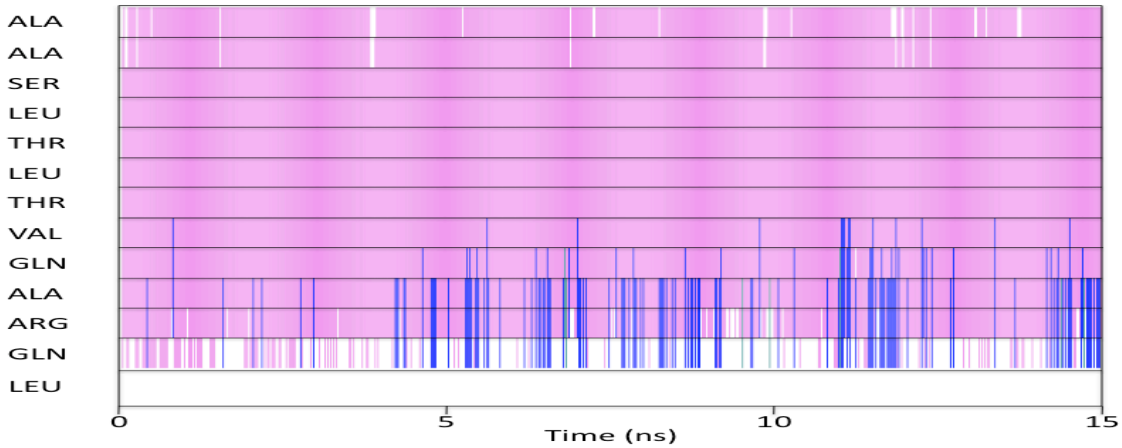


Figure 4.5: Secondary structure of each residue of the initially 'helical' $gp41_{532-544}$ as a function of time. Pink corresponds to α-helix, blue is 3_{10}- helix, green is turn and white corresponds to unordered.

This simulation therefore appeared to suggest that the helical conformation was indeed a stable structural motif for this peptide, however such secondary structure was not found experimentally under aqueous conditions. As highlighted in *Section 3.4.1.3*, it has been suggested that classical force fields over express helical content and that different force fields may differ widely in the regions of dihedral space sampled [32]. Furthermore, from the previous investigation into starting structure bias introduced in *Section 3.4.1.2*, CHARMM22 appeared to introduce an "effective temperature shift" and trap the simulation within a helical well on the energy landscape [93]. The simulation of $gp41_{532-544}$ starting from the helical conformation may therefore have experienced the same obstacle, in that the AMBER-ff03 force field was not capable of fully

sampling the energy landscape. However in the investigation of $gp41_{659-671}$ in *Chapter 3*, the AMBER-ff03 force field was found to yield a significant improvement over the CHARMM22 parameterisation, therefore it was important to consider simulating the peptide from a different starting structure.

In preparation for simulating a dimeric system consisting of $gp41_{532-544:659-671}$, it was decided to simulate $gp41_{532-544}$ starting from the same initial configuration as used previously for $gp41_{659-671}$ - the unordered structure reported by Barbato *et a.* [2]. Firstly, this would allow for a direct like-for-like comparison between the two peptide simulations, ensuring that differences in the simulated structures were a result of the influence that the primary amino-acid sequence had on the folded conformations and not as a result of a bias introduced by the starting structure. Secondly, it would allow for the structures predicted from an initially unordered dimeric system, to be compared directly to the individual monomeric counterparts, therefore revealing whether the presence of one peptide influenced the conformational space sampled by the other.

The simulation of the initially unordered $gp41_{532-544}$ monomer revealed the secondary structure of the peptide to be comprised of two helical regions, with a kink localised on the Thr_{538} residue which appeared to act as a hinge between these two regions (Fig. 4.6a). This was confirmed by analysis of the backbone torsional angles over the last 5ns of the simulation (Fig. 4.6b) which revealed that the main feature of the Ramachandran plot was a compact peak located with the right-handed α-helical region of dihedral space, with only the Thr_{538} residue presenting a peak outwith this region (located within the left-handed α-helical region).

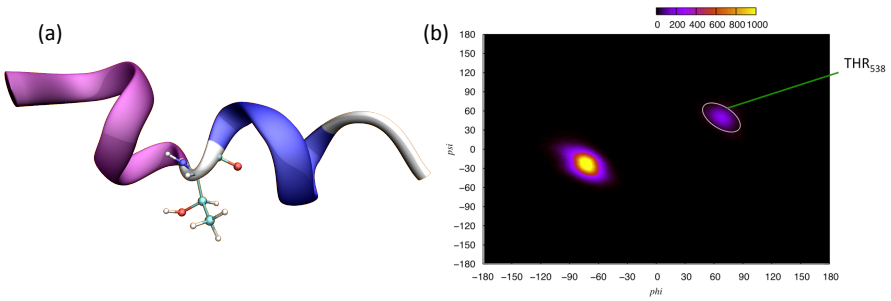


Figure 4.6: (a) Structure after 15ns of initially 'unordered' simulation of $gp41_{532-544}$, revealing kinked helix. (b) Ramachandran plot of the simulation over the last 5ns, identifying Thr_{538} as the only amino acid outside right-handed α-helical region.

In order to probe the stability of the two helical regions over the course of the simulation, the STRIDE secondary structure algorithm was employed to follow the secondary structure of

each residue with respect to time (Fig. 4.7). This revealed that the structure after the equilibration phase was dominated by a mixture of turn-like and 3_{10}-helix motifs. Within 1ns of the simulation, the structure over the sequence Ala_{532}-Leu_{537} was dominated by an α-helical motif which persisted over the remainder of the 15ns simulation. The C-terminal was found to be more dynamic, displaying a mixture of turn, 3_{10}-, and α-helical content. While this region was initially dominated by turn-like motifs, the sequence Val_{539}-Gln_{543} was found to adopt a primarily 3_{10}-helical structure as the simulation progressed. In comparison to the simulation starting from the initial helical conformation, both parameterisations predict the peptide to adopt a predominantly helical structure (a structure not observed experimentally) characterised by an α-helical N-terminus and a C-terminal defined by a mixture of turn, α- and 3_{10}-helix (the exact ratio being dependent on the initial configuration).

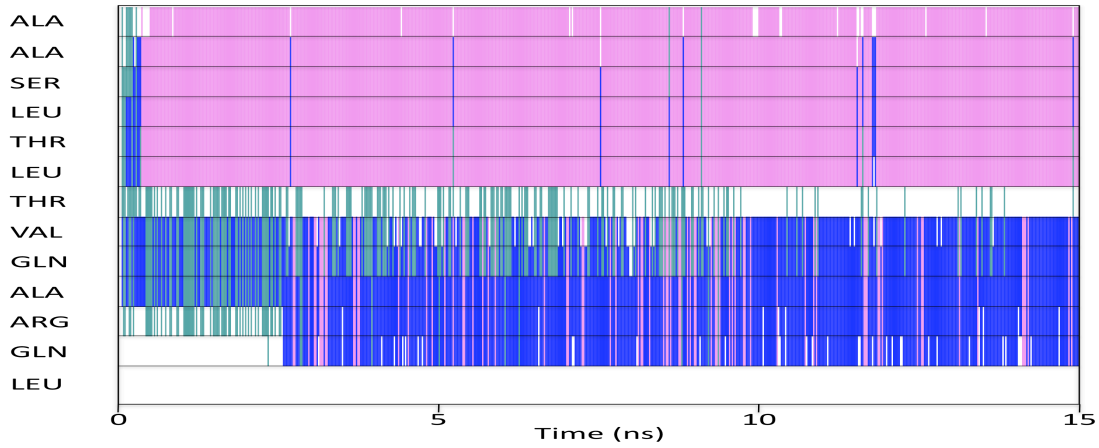


Figure 4.7: Secondary structure of each residue of the initially 'unordered' $gp41_{532-544}$ as a function of time. Pink corresponds to α-helix, blue is 3_{10}- helix, green is turn and white corresponds to unordered.

Having started from the same initial structure as the simulation of $gp41_{659-671}$ (Fig. 3.3), it was therefore reassuring that the simulation predicted different conformations than those found in the previous model and did not fold into the same structure regardless of the major differences in primary sequence. Furthermore, as the simulation of $gp41_{659-671}$ predicted a rigid helix over Glu_{662}-Ser_{668}, the dynamic nature of the C-terminal of $gp41_{532-544}$ would indicate that this FPPR peptide should exhibit a better conformational responsiveness than it's MPER counterpart.

With a view to probing the various structural motifs that $gp41_{532-544}$ could explore, it was important to study the system under conditions mimicking the environmental changes implicated

in the fusion process, e.g. dielectric constant fluctuations and interfacial interactions [81, 152]. As such, it was decided to subject this peptide to the same organic solvents and surfactant-containing buffers previously used in conjunction with $gp41_{659-671}$, i.e. TFE, SDS, acetonitrile, and methanol.

As with the MPER $gp41_{659-671}$, titration of $gp41_{532-544}$ with TFE resulted in a shift in the conformer population towards higher percentages of helical structures (Fig. 4.8a). This was indicated by the formation of a band of positive ellipticity centred around 190 nm, the shift of the negative peak from 197 nm to 206 nm and the appearance of a small shoulder at around 224 nm. The stabilisation of such a motif was not unexpected as the simulation predicted a predilection for a helical conformation, and TFE is known to stabilise the secondary structure for which the amino-acid sequence has a propensity [114]. Other mixed-solvent conditions (50% Methanol and 50% Acetonitrile, v/v) similarly resulted in the ensemble of peptide conformations shifting towards a higher helical content (Fig. 4.8b), with the changes in the CD spectra being qualitatively similar to those observed in TFE. The promotion of intramolecular hydrogen bonding within the peptide, resulting in the $gp41_{532-544}$ folding into a helical conformation, was no doubt a result of the lower dielectric constant of these solvents. Due to these mixed solvents being mimetics of the extended regions of a membrane [81], it can be concluded that similarly to its MPER counterpart, the secondary structure of the FPPR $gp41_{532-544}$ is more ordered in its native state near a membrane surface.

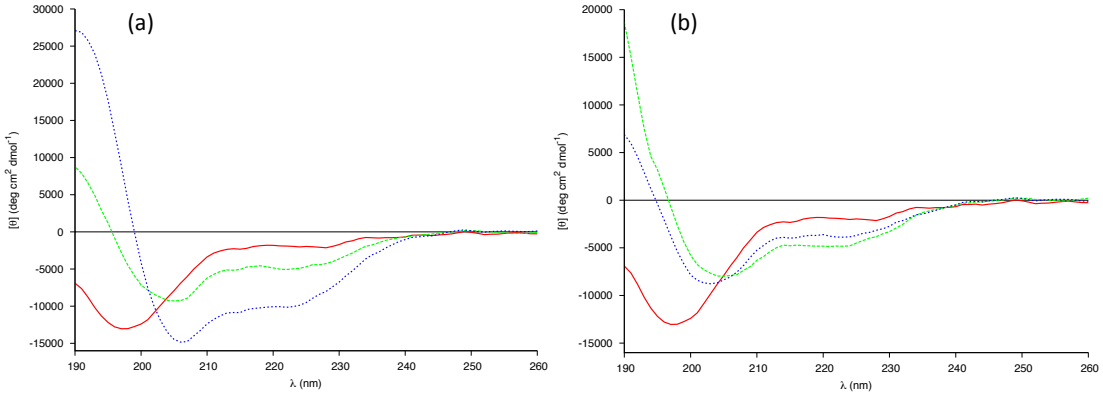


Figure 4.8: Far-UV CD spectra of 0.04mM $gp41_{532-544}$ in 10mM phosphate buffer (red) and (a) 20% (green) & 60% (blue) TFE, and (b) 50% Methanol (green) & 50% Acetonitrile (blue).

In *Section 3.4.3*, the cumulative partial denaturation effects of a membrane surface were investigated utilising micellar SDS, providing an anionic membrane-mimicking environment. As

the FPPR stretch should be located near the host cell membrane, it would be more appropriate to investigate the structural changes in a zwitterionic (mammalian) membrane-mimicking environment rather than SDS. However, it was important to be consistent in the samples used for conformational sampling, while at the same time it was interesting to probe electrostatic implications; variations in the pH were exploited to probe the electrostatic interactions between $gp41_{659-671}$ and the negatively charged head groups of SDS, therefore it was of considerable interest to probe such interactions with the positively charged species $gp41_{532-544}$.

Far-UV CD spectra of $gp41_{532-544}$ were recorded in the presence of both non-micellar and micellar SDS (Fig. 4.9). The spectra revealed a concentration-dependent folding in micellar SDS (12mM), with the peptide folding into a helical structure at concentrations of 40 μM, while it was found to adopt a β-structure at higher concentrations of 240 μM; concurrently, at non-micellar concentrations (3 mM), $gp41_{532-544}$ folded into a β-sheet structure at both concentrations. This was inferred from the formation of a band of positive ellipticity at 195nm plus a negative band between 215-220nm, and a positive band centred at 190nm plus two bands of negative ellipticity at 205nm & 220nm, for the β-sheet and α-helical structures respectively. Such a concentration-dependence polymorphism was not observed in the $gp41_{659-671}$ peptide, suggesting that the folding of the $gp41_{532-544}$ peptide was controlled by the surface of the micelle.

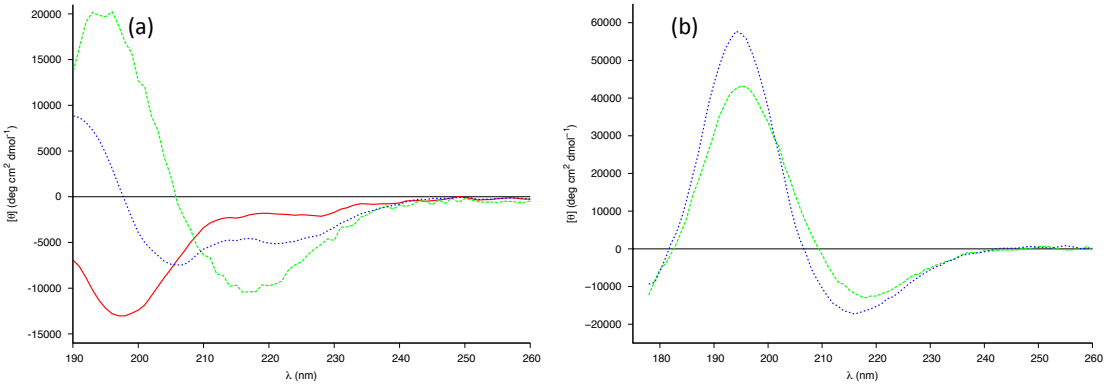


Figure 4.9: Far-UV CD spectra of (a) 0.04mM and (b) 0.24mM $gp41_{532-544}$ in 10mM phosphate buffer (red), 3mM (green) and 12mM (blue).

With regards to $gp41_{659-671}$, the stabilisation of a helical conformation in organic solvents was attributed to the solvent locking the peptide within a hydrophobic matrix of intramolecular hydrogen bonds (*Section 3.4.2*). In the case of $gp41_{532-544}$, it would appear that such a confirmation could be compromised by non-specific charge interactions between the positively charged

$gp41_{532-544}$ and the anionic SDS monomers (i.e. at concentrations below cmc). At concentrations above the cmc, SDS micelles provide an extensive surface support (resembling a bulk organic phase), and by lowering the local dielectric constant, reduces hydrogen bonding between the micellar surfaces and peptide favouring structures that are stabilised intramolecularly (i.e. helices). The concentration dependence would appear to indicate that upon increasing peptide concentration, peptide-SDS electrostatic interactions recover β-sheet conformations, predominantly stabilised by peptide-peptide and peptide-solvent intermolecular hydrogen bonding.

4.5.2 Monomers at membrane environments

In order to asses the ability of the host-cell lipid environment to induce folding, both $gp41_{532-544}$ and $gp41_{659-671}$ were investigated in the presence of zwitterionic membranes. None of the acquired far-UV CD spectra displayed any appreciable secondary structure at either neutral or acidic pH levels (Fig. 4.10). The lack of folding in the $gp41_{532-544}$ stretch was inline with the current fusion models which describe the FP as inserting into the cellular membrane without supporting or accompanying FPPR-membrane interactions [142–144].

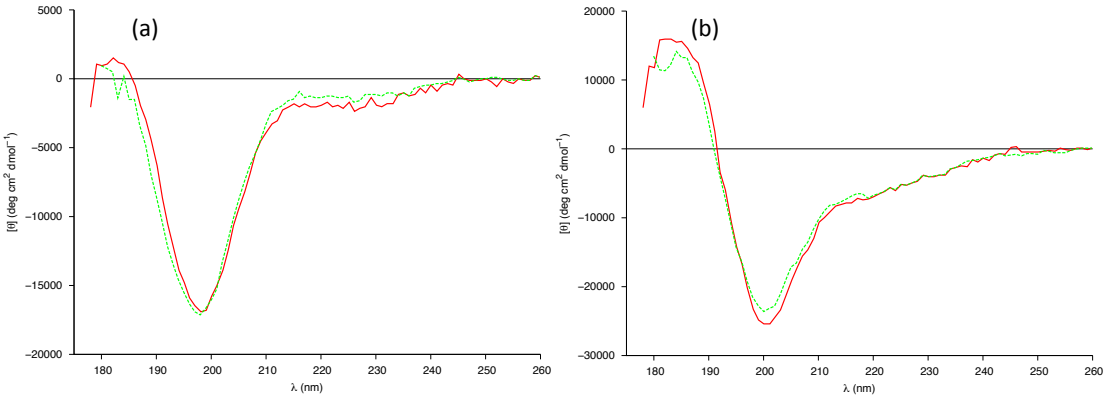


Figure 4.10: Far-UV CD spectra of (a) 0.2mM $gp41_{659-671}$ and (b) 0.2mM $gp41_{532-544}$, in the presence of 1.6mM zwitterionic membranes and 10mM PB (red) or H_2O, pH 3 (green).

The native environment of $gp41_{659-671}$ in both the pre-fusion and hemifusion phases was subsequently emulated utilising anionic vesicles; in both phases, the $gp41_{659-671}$ stretch is in the proximity of the viral membrane where it has been reported to remain unfolded [158, 159]. No folding was observed in the recorded far-UV CD spectra of $gp41_{659-671}$ in anionic membranes, irrespective of peptide concentration or pH (Fig. 4.11a). The partial structuring of this stretch in SDS micelles was a result of pH-induced conformational locking of the $gp41_{666-670}$

Trp span, therefore changes in the microenvironment of the span were probed using intrinsic fluorescence in order to probe whether the locking was a pre-requisite for the hydrophobic interactions with the anionic membrane. As there were no observed shifts in the fluorescence spectra for gp41$_{659-671}$ with and without lipids (Fig. 4.11b), this indicated that there were negligible changes in the microenvironment and there were no folding-related interactions of the stretch. This was supported by linear dichroism (LD) measurements conducted on this peptide in the presence of anionic vesicles [160], which suggested that the peptide adopted an unfolded structure and lay parallel with respect to the membrane surface.

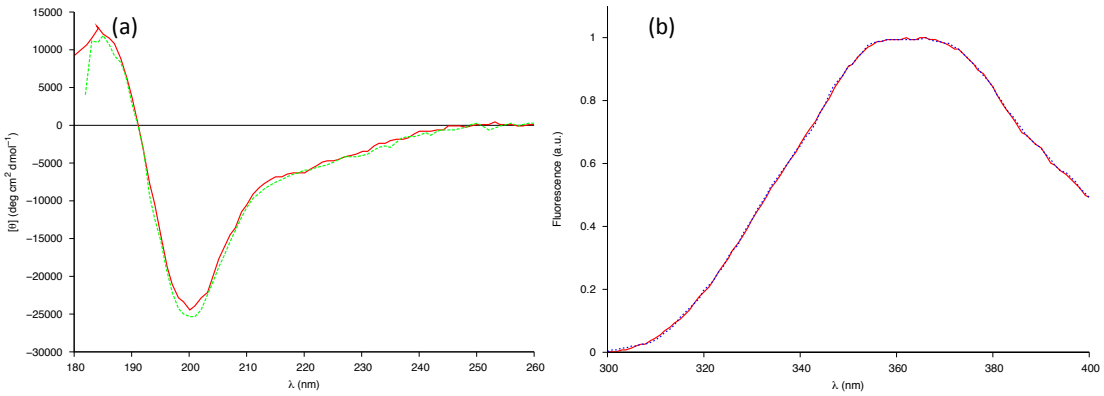


Figure 4.11: (a) Far-UV CD spectra of 0.19mM gp41$_{659-671}$ in the presence of 1.6mM anionic membranes and 10mM PB, pH 7.5 (red) or H_2O, pH 3 (green). (b) Normalised Trp fluorescence of 0.04mM gp41$_{659-671}$ (red) and in the presence of anionic membranes (blue).

In contrast to the observations made in neutral membranes, far-UV CD spectra of gp41$_{532-544}$ in anionic vesicles were characteristic of a β-sheet structure. This folding was not found to be dependent on the concentration of the protein (Fig. 4.12a) as was observed for the peptide in the presence of micellar SDS (Fig. 4.9). However, such folding was found to be pH-dependent, with a transition to a random coil being observed upon lowering the pH (Fig. 4.12b). As this resulted in charge quenching of the anionic membrane, such a transition implied that electrostatic interactions were decisive for the membrane-mediated folding of gp41$_{532-544}$.

Far-UV CD spectra of increasing amounts of the membrane solution (Fig. 4.13) revealed that the folding was complete at a ratio of 10:1, at which point gp41$_{532-544}$ was essentially 100% membrane-bound. The folding of a β-sheet structure was supported by LD measurements conducted on this peptide in the presence of anionic vesicles, with the spectra indicating a β-sheet structure lying on the membrane surface [160], and the preferential folding is likely promoted by

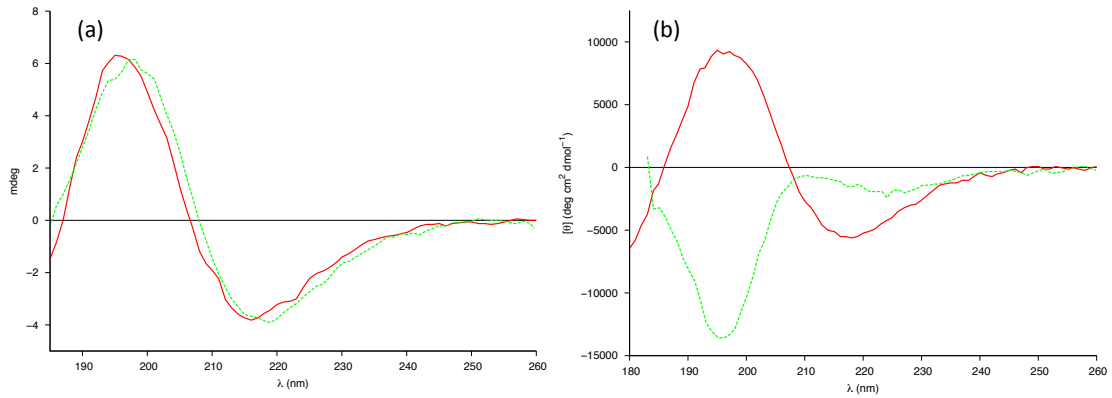


Figure 4.12: Far-UV CD spectra of (a) 0.04mM gp41$_{532-544}$ (red) & 0.4mM gp41$_{532-544}$ (green), and, (b) 0.24mM gp41$_{532-544}$ in 10mM phosphate buffer, pH 7.5 (red) or H_2O, pH 3 (green) - all in the presence of 1.6mM anionic membranes.

the gp41$_{535-540}$ having a high β-structure propensity (Tab. A.4).

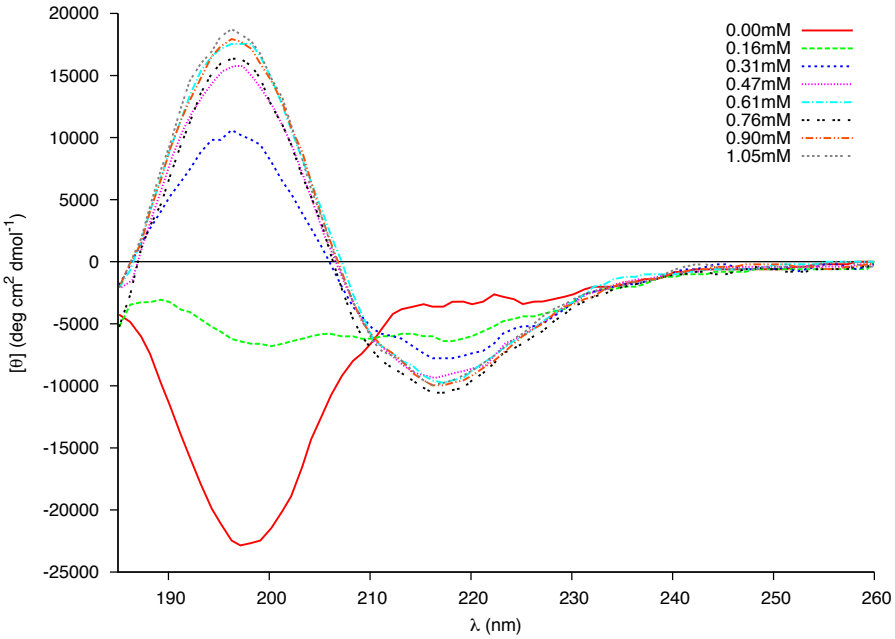


Figure 4.13: Far-UV CD spectra of 0.04mM gp41$_{532-544}$ in the presence of increasing concentrations of anionic membranes.

The stability of gp41$_{532-544}$ and gp41$_{659-671}$ was investigated by following the thermal transitions of these peptides in the presence of anionic membranes. As was expected, far-UV CD spectra of gp41$_{659-671}$ in such an environment were characteristic of unfolded structures over several temperatures (Fig. 4.14a). However, far-UV CD spectra of gp41$_{532-544}$ and anionic vesicles suggested a stabilisation of the β-structure at temperatures in excess of 40°C; characterised by an increase in intensity for the band of positivity ellipticity at 196 nm, and a shift in

the negative band to 218 nm (Fig. 4.14b). This was attributed to extended surface contacts as a result of increased lipid migration, which is in agreement with the temperature-dependent lipid motion from a solid (gel) phase at 20°C through a rippled phase at >40°C, into a fluidic phase at 60°C [161]. A caveat of such lipid migration, is that an increase in lipid diffusion must be matched with a sufficient peptide concentration in order to provide ample coverage and binding. This is exemplified by $gp41_{532-544}$ at lower concentrations, which underwent thermal unfolding (Fig. 4.14c). As it is suggested that the fluidic phase is the best mimetic of biological membranes, it stands to reason that the observed increase in folding supports a cooperative and responsive mode of $gp41_{532-544}$ binding.

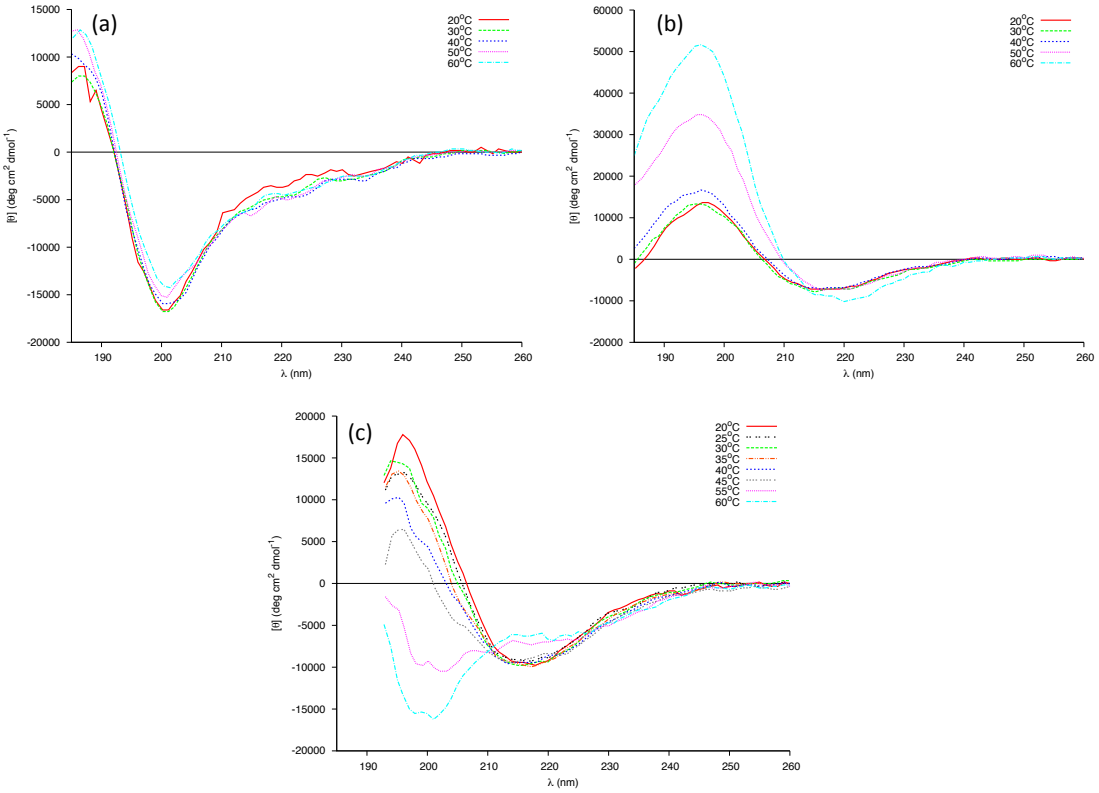


Figure 4.14: Far-UV CD spectra of (a) 0.04mM $gp41_{659-671}$, (b) 0.037mM $gp41_{532-544}$ and (c) 0.008mM $gp41_{532-544}$, in the presence of 1.6mM anionic membranes over a range of temperatures.

4.5.3 Dimer System

In order to investigate whether the $gp41_{532-544}$ and $gp41_{659-671}$ sequences fold cooperatively with one another, far-UV CD spectra of an equimolar mixture of $gp41_{532-544:659-671}$ in 10mM

PB were recorded. Utilising the quantitative nature of CD, this spectra was directly compared to a 'theoretical' linear combination of the individual monomeric spectra for gp41$_{532-544}$ and gp41$_{659-671}$ (Eq. 4.1). Such a comparison would highlight any structural variation resulting from the cooperative folding of the dimeric system. However, as was expected, no appreciable structuring was observed under aqueous conditions, with the shape of both spectra being representative of an unordered confirmation (Fig. 4.15a). Furthermore, such a comparison was extended to include the equimolar mixture of gp41$_{532-544:659-671}$ in 10mM PB and either 20% TFE, 50% Acetonitrile, or 12mM SDS (Fig. 4.15b-d). In each condition tested, the linear combination of the monomeric spectra was revealed to be essentially identical to the far-UV CD spectra recorded for the equimolar mixture. This not only supported the finding that these sequences did not fold cooperatively in the absence of membranes, but also revealed that induced pre-folding (arising from the solvents used; Acetonitrile, TFE, and SDS) was not sufficient to promote the co-ordering of the stretches.

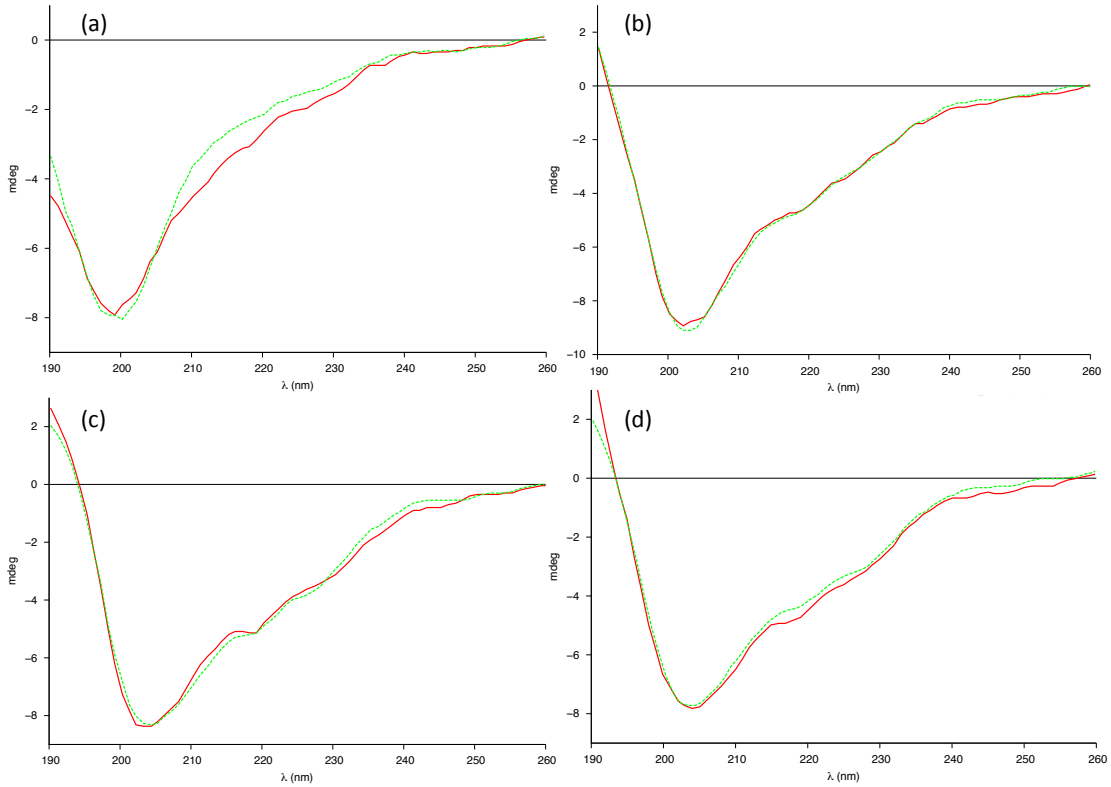


Figure 4.15: Far-UV CD spectra of equimolar gp41$_{532-544:659-671}$ (red) compared to the linear combination of the monomeric spectra (green) in (a) 10mM phosphate buffer, (b) 50% Acetonitrile, (c) 12mM SDS, and (d) 20% TFE.

Having simulated the individual monomers from a predominantly unordered confirmation, it was therefore of interest to investigate whether the final structures observed for these simulations would be reproduced in a dimeric system, or whether the presence of one stretch of amino acids would perturb the final structure of the other. The starting point for the dimer simulation of $gp41_{532-544:659-671}$ was therefore an anti-parallel arrangement of the unstructured stretches of $gp41_{532-544}$ and $gp41_{659-671}$ used in the corresponding monomer simulations. Analysis of the simulation utilising the STRIDE secondary structure algorithm (Fig. 4.16) revealed that, interestingly, the secondary structure features of the monomeric $gp41_{532-544}$ and $gp41_{659-671}$ appeared to be retained in the dimeric $gp41_{532-544:659-671}$; $gp41_{532-544}$ was still defined by two helical regions spanning amino-acids Ala_{532}-Leu_{537} and Val_{539}-Gln_{543}, while $gp41_{659-671}$ was characterised by a helix between Glu_{662}-Ser_{668}. This thereby appeared to support the results obtained experimentally and suggested that under aqueous conditions, the individual stretches did not fold cooperatively.

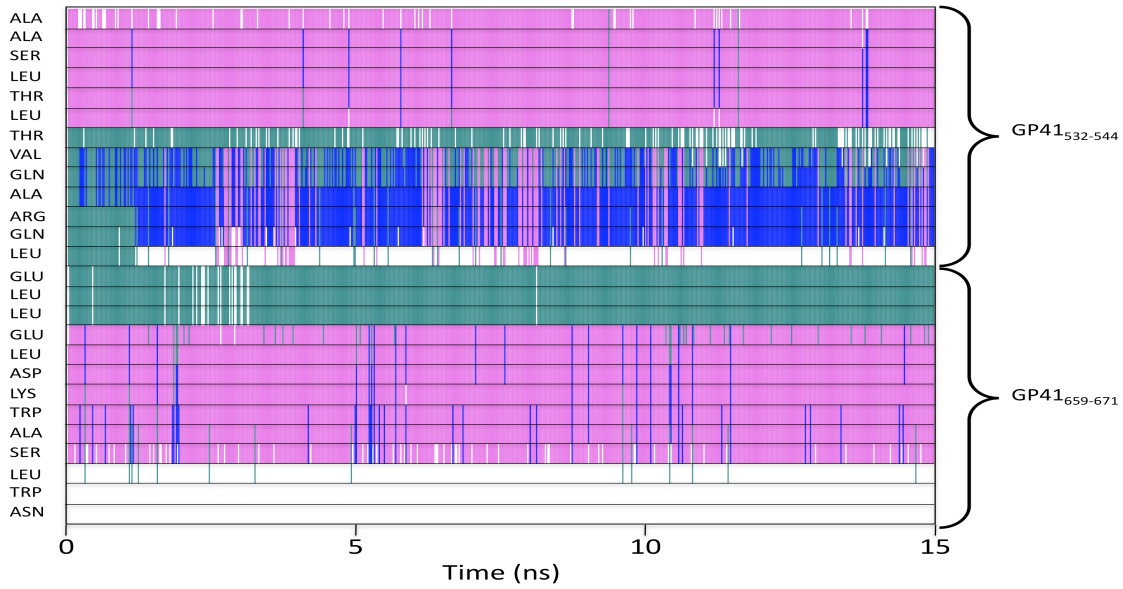


Figure 4.16: Secondary structure of each residue of the dimeric $gp41_{532-544:659-671}$ as a function of time. Pink corresponds to α-helix, blue is 3_{10}- helix, green is turn and white corresponds to unordered. Simulation does not reveal any major changes to the main secondary structural features.

As the dimeric system under solvent conditions did not reveal any notable interactions between the two stretches, it was important to probe if these systems influenced one another in a lipidic environment. In comparing the far-UV CD spectra of $gp41_{532-544:659-671}$ in the presence of

1.6mM DLPC liposomes and 10mM PB, to the linear combination of the non-interacting individual monomers under identical conditions (Fig. 4.17), it was apparent that the dimeric system did not display an appreciable secondary structure and revealed no interactions at zwitterionic membrane surfaces. Synergetic interactions between the FPPR and MPER regions, both implicated in the fusion process [133, 162], would likely transpire in mixed lipid environments as they are brought together upon formation of the 6HB. Therefore, one would expect an appreciable co-operative folding between $gp41_{532-544}$ and $gp41_{659-671}$ in the presence of anionic membranes. From the monomeric stretches, it was apparent that the activity of $gp41_{659-671}$ (which remained unfolded in the anionic environment) would be determined by the responsive membrane binding of its $gp41_{532-544}$ counterpart which folded into a β-sheet structure.

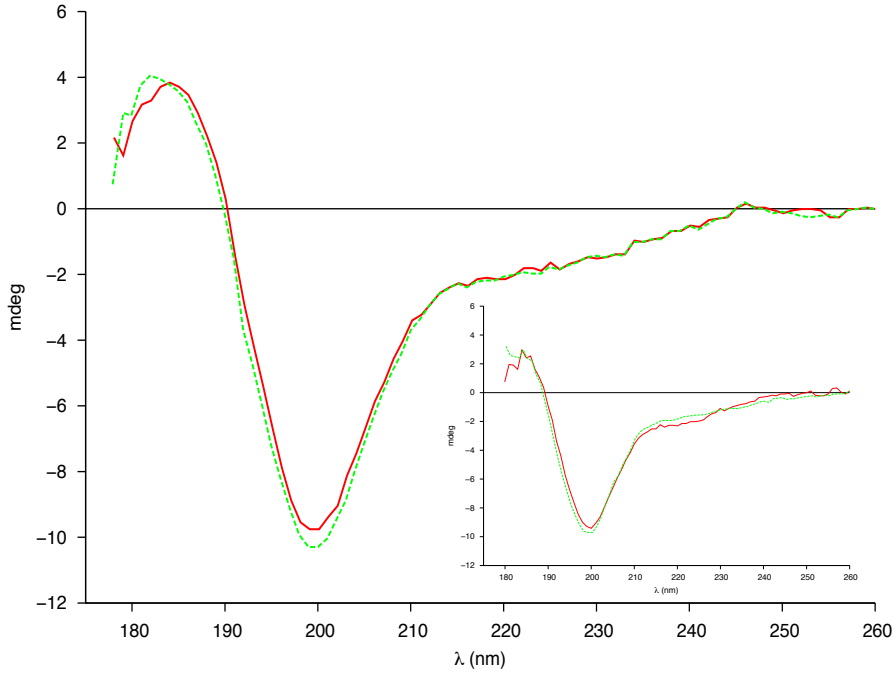


Figure 4.17: Far-UV CD spectra of dimeric 0.04mM $gp41_{532-544:659-671}$ (red) compared to the linear combination of the monomeric spectra (green) in the presence of 1.6mM zwitterionic membranes and 10mM phosphate buffer, pH 7. Inset: in the presence of 1.6mM zwitterionic membranes and H_2O, pH 3.

The equimolar mixture of $gp41_{532-544}$ and $gp41_{659-671}$ ($gp41_{532-544:659-671}$) was titrated with anionic membranes. Far-UV CD spectra revealed a series of conformation responses in the mixture as a function of increasing peptide-lipid rations. At a ratio of 1:4 (corresponding to a peptide and lipid concentration of 0.04mM and 0.16mM respectively) a transition to a helical motif was observed, indicated by the emergence of a band of positive ellipticity at 190nm

and two negative bands at 202nm and 217nm (Fig. 4.18b). Upon increasing the lipid concentration to 0.9mM, resulting in a ratio of approximately 1:20, the recorded CD spectra suggested a conformational shift towards β-type conformers (Fig. 4.18c); however, the positive peak near 190nm is usually characteristic of helical structure. These spectra were subsequently compared to the linearly combined spectra of the monomeric data at appropriate concentrations, which assumed that there were no perturbations to the structure upon mixing; obtained from titrating the individual peptides with anionic membranes. As there were no differences between the linear combination and acquired spectra of the dimeric gp41$_{532-544:659-671}$ system in lipid-free media, it would appear that any potential binding interplay between the two stretches ensues as a result of saturating peptide concentrations onto membrane surfaces at the 1:4 ratio.

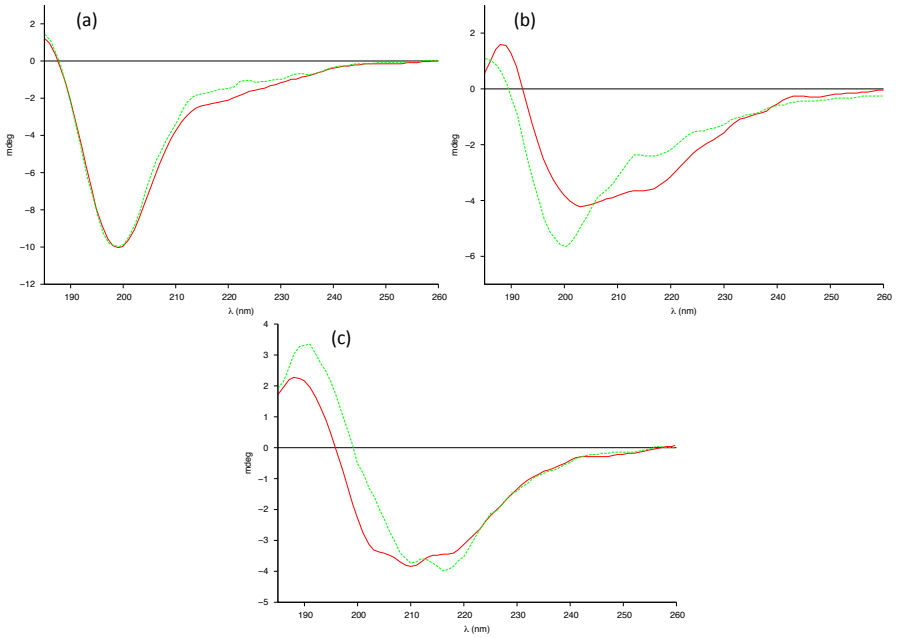


Figure 4.18: Far-UV CD spectra of 0.04mM gp41$_{532-544:659-671}$ (red) compared to the linear combination of the monomeric spectra (green) in the presence of (a) 10mM phosphate buffer (b) 0.16mM anionic membranes, and (c) 0.9mM anionic membranes.

Within the 6HB, both stretches fold into α-helices as terminal parts of their corresponding heptad repeat regions, and can therefore be expected to exhibit preferential helical folding. However as observed in *Section 4.5.2*, this was not true for the monomeric systems in the presence of either lipidic system (i.e. DLPC or DLPC/DLPG), or for the dimeric mixture of gp41$_{532-544:659-671}$ in the absence of such membranes; in fact, gp41$_{532-544}$ was found to adopt β-sheet membrane-induced conformations, whereas gp41$_{659-671}$ was not found to bind to mem-

branes in a folding-dependent manner. It would therefore appear that the dimer would only able to emulate the native tertiary contacts of the regions in the 6HB while in the presence of anionic membranes.

Fluorescence spectra of $gp41_{532-544:659-671}$ were recorded for any changes in intrinsic fluorescence as compared with the results for $gp41_{659-671}$. As no changes were observed, this suggested that the $gp41_{666-670}$ Trp span did not interact with the membranes the way it did in SDS micelles (Fig. 4.19): the span may preferentially interact within $gp41_{532-544:659-671}$. In support of this, LD measurements conducted on the $gp41_{532-544:659-671}$ dimer system suggested that the tryptophans were not involved in the dimers interactions with the membrane [160]. Furthermore, the LD results suggested the dimer system exhibited multiple binding modes consisting of a predominantly perpendicular orientation of helices to the membrane surface, and residual β-structure parallel to the surface. As the collected LD data could not rule out that one stretch was orientated parallel to the membrane surface, whilst the other was perpendicular, it was important to return to the simulation to gain further insight into the orientation of the stretches within the dimer.

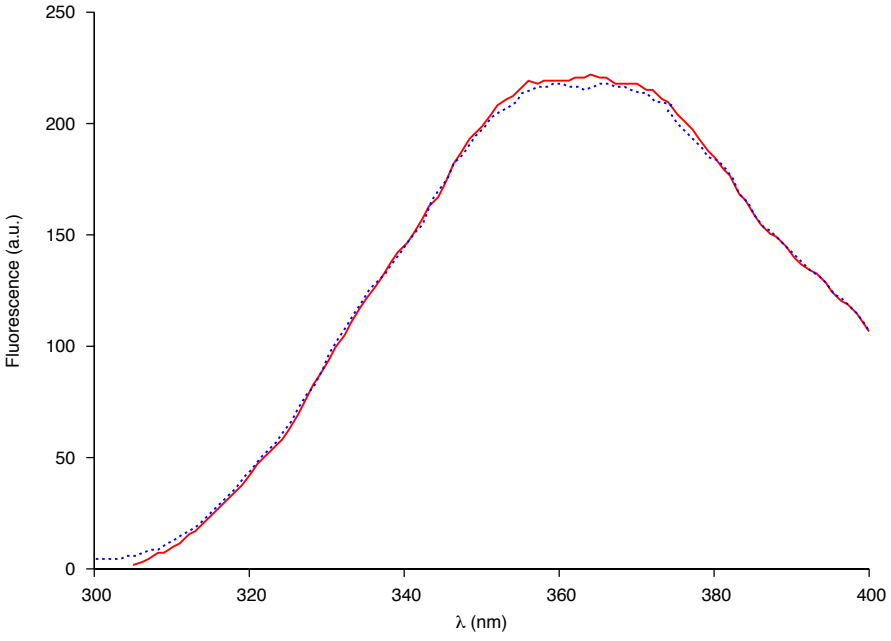


Figure 4.19: Fluorescence spectra of 0.02mM $gp41_{659-671}$ (red) and 0.04mM $gp41_{532-544:659-671}$ (blue) in 10mM PB and in the presence of 1.6mM anionic liposomes.

As mentioned in *Section 4.4.5*, the starting point for the simulation of $gp41_{532-544:659-671}$ was an anti-parallel arrangement of the unstructured stretches of $gp41_{532-544}$ and $gp41_{659-671}$.

Following an equilibration phase, visualisation of the simulation revealed that the anti-parallel arrangement of the monomers was maintained within the first 5ns (Fig.. 4.20a), whereafter the peptides arranged in a perpendicular fashion, with the C-terminal of $gp41_{659-671}$ orientated towards $gp41_{532-544}$ (Fig.. 4.20b). This transition appeared to be driven by the residues Leu_{535}, Trp_{666}, Leu_{669}, and Trp_{670}, forming a network of hydrogen bonds within a hydrophobic interface; during the transition, the residues Leu_{535} and Trp_{670} appeared to form the dominating interaction (Fig.. 4.20c).

As the simulation evolved up to 10ns (of the 15ns production run), the Trp_{666} residue relocated from the hydrophobic core at the expense of stabilising interactions with Leu_{663} (Fig. 4.21a) before then folding back into the hydrophobic core. In the latter stages of the simulation, the two α-helical regions of $gp41_{532-544}$ and $gp41_{659-671}$ appeared to be aligned in parallel before drifting apart by 12.5ns into the simulation (Fig.. 4.21b), with Trp_{666} and Trp_{670} forming close contacts with Leu_{663} and Leu_{669} respectively. The simulation of the dimer predicts that $gp41_{532-544:659-671}$ is indeed a highly dynamic structure with an intrinsic ability to rearrange. Different structural arrangements of the pair appear to be controlled by hydrophobic inter- and intra-peptide interactions, which are primarily mediated by the Trp residues of $gp41_{659-671}$. Complementarily, the kinked helix of $gp41_{532-544}$ can provide conformational flexibility necessary to meet geometric requirements of membrane rearrangements implicated in phase transitions.

4.6 Conclusions

The membrane proximal regions of gp41 have been shown to interact cooperatively, exclusively in membrane-bound environments. Specifically, the binding interplay between the two highly conserved peptide stretches $gp41_{532-544}$ and $gp41_{659-671}$ (of FPPR and MPER, respectively) resulted in the formation of helical structure in the presence of anionic membranes; neither of the individual monomers in the presence of anionic or zwitterionic membranes displayed such structural propensity. Furthermore, these peptides have been shown to fold into helices without the presence of the HR1 and HR2 stretches. This autonomy suggests that $gp41_{532-544:659-671}$ can serve as a functional bridge between the FP, TMR and 6HB, thereby supporting the lipid-

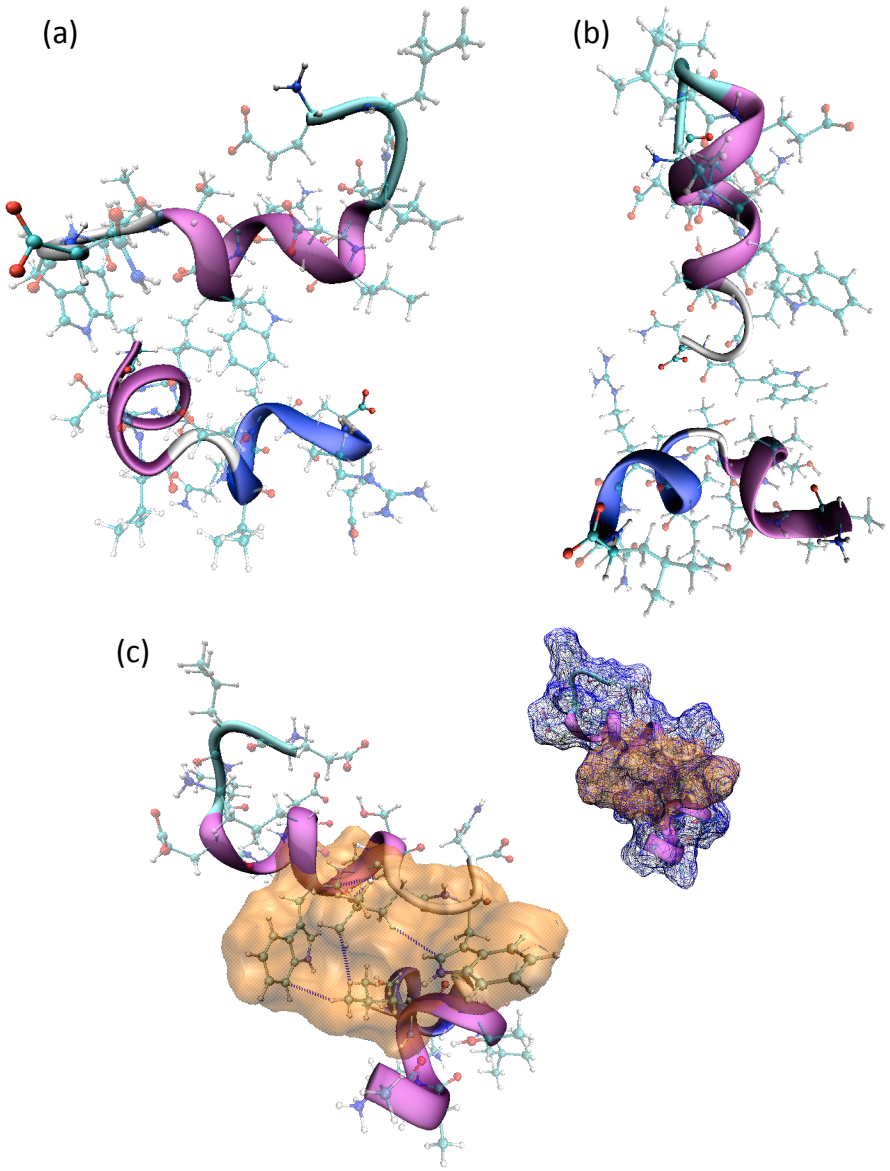


Figure 4.20: Visualisation of the dimeric $gp41_{532-544:659-671}$ simulation. (a) after the equilibration phase revealing that the initial anti-parallel arrangement was retained. (b) after 5ns revealing the stretches in a perpendicular arrangement. (c) a hydrophobic interface (orange) comprising Leu_{535}, Trp_{666}, Leu_{669}, and Trp670. Inset: a complete molecular surface (blue) with the interface in the centre.

mixing phase and therefore implies that it may possess a regulatory function [156, 163]. As isolated $gp41_{659-671}$ was found to posses negligible membrane interaction while $gp41_{532-544}$ was found to display distinct folding-mediated binding [160], it is reasonable to assume that it is the FPPR stretch that acts as the primary regulator of the MPER-FPPR interactions in the

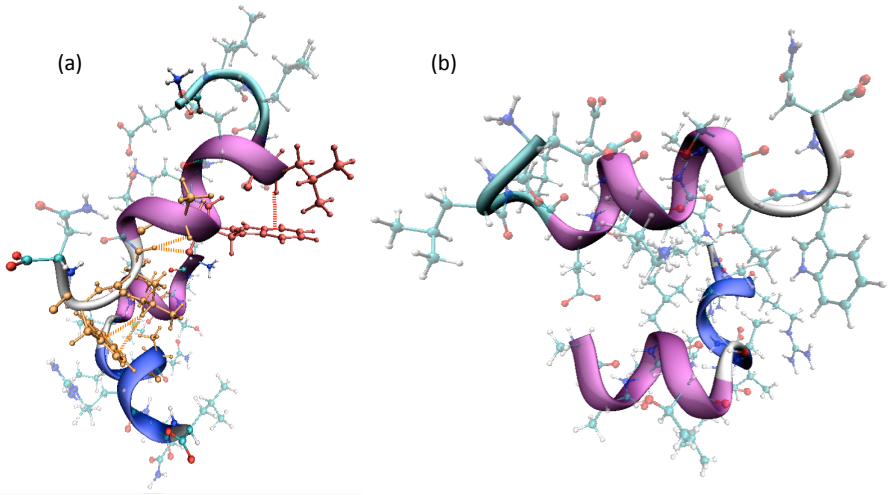


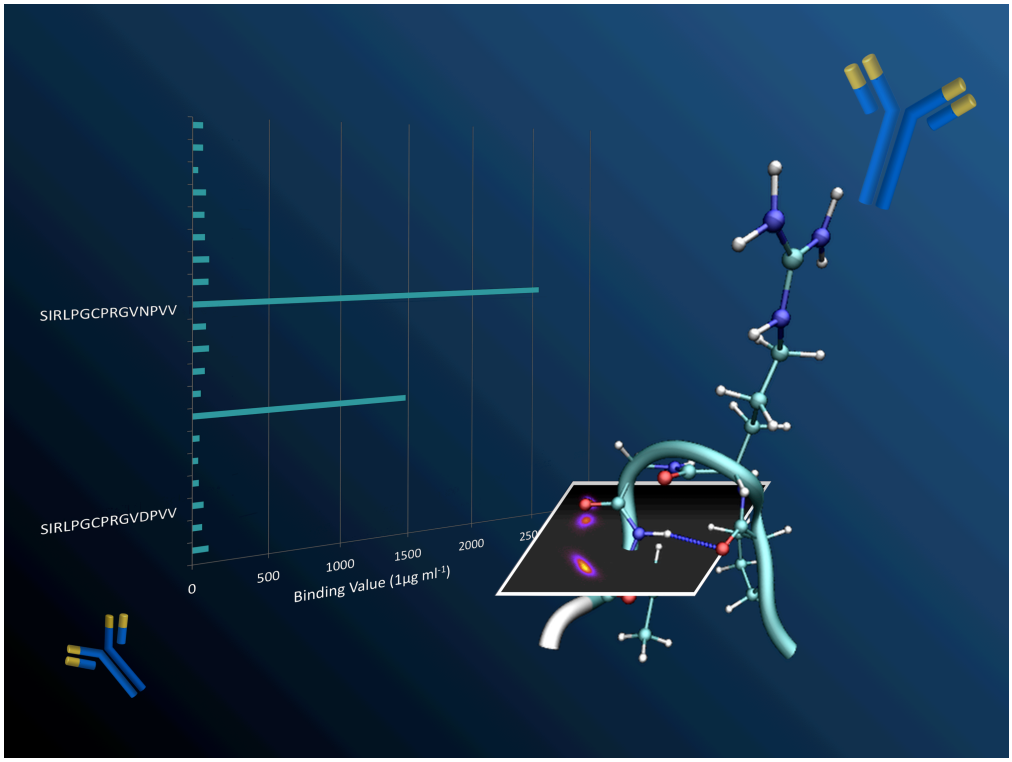
Figure 4.21: Snapshot of the dimeric $gp41_{532-544:659-671}$ simulation. (a) Trp_{666} residue leaving the hydrophobic interface (orange) to interact with Leu_{663} (red). (b) α-helical regions of $gp41_{532-544}$ and $gp41_{659-671}$ aligned in parallel.

hemifusion phase [107, 164, 165].

The results of this study suggest that in the Env, MPER and FPPR are co-aligned without the ability to fold or assemble. In the pre-fusion phase, MPER displays negligible interacts with the viral membrane, while the FPPR does not interact with the cell membrane at all. However, in the hemifusion phase (characterised by the changed lipid environment), the interactions of the FPPR responds to the proximity of the viral membrane, which in turn triggers the responsive binding of the MPER.

5

Effects of point mutation: Antibody recognition of the epitope hCGβ_{66-80}.



5.1 Introduction

The glycoprotein human chorionic gonadotropin (hCG), is a hetero-dimer consisting of an α-chain in non-covalent association with a β-chain (Fig. 5.1). Essential in pregnancy and detectable within a few days of fertilisation, hCG has become one of the most frequently assayed hormones, and simple, one-step, antibody-based measurements of hCG are now standard in pregnancy testing. At this early stage of pregnancy, the trophoblast cells of the pre-implantation embryo produce a hyperglycosylated form of the hormone, which drives embryonic implantation in the uterine wall [166]. Once successful implantation has been achieved, hCG production is controlled by syncytiotrophoblast cells of the placenta, and its continued action is required to promote and maintain the secretion of progesterone from the corpus luteum, which is essential in maintaining the pregnancy. In other applications, therapeutically, injected hCG is the key to *in-vitro* fertilisation, where it is used to trigger ovulation and, in males, it is used to promote testosterone production.

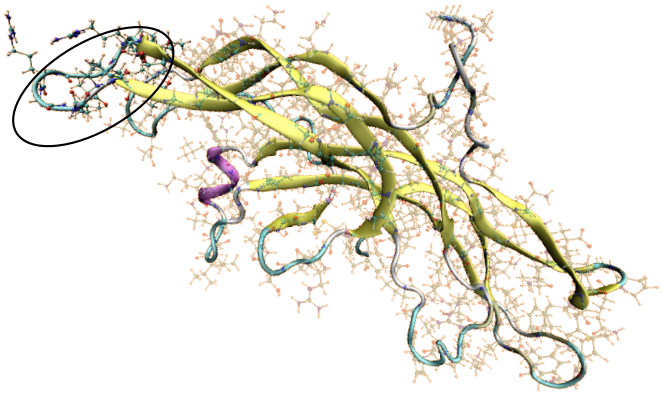


Figure 5.1: Crystal structure of hCG [4], hCGβ_{66-80} is circled.

The role of hCG as a tumour marker in a variety of cancers is of great clinical significance [167, 168], and quantitative, specific hCG tests are important in the diagnosis and management of hCG-secreting cancers [169]. However the design criteria of hCG cancer assays differ from those required for pregnancy tests. Specifically, different sensitivity ranges are involved and, for oncology applications, the antibodies used must be able to detect various modified forms of hCG with equal affinity [170]. The role of hCG assays in the diagnosis and management of malignant trophoblastic disease (choriocarcinoma) is one of the great success stories of onco-

logy [171, 172]; it is the ideal tumour marker - always present when choriocarcinoma cells exist, and in quantities directly related to the number of those cells. The correct use of hCG assays in combination with appropriate therapy has led to a survival rate approaching 100% in this otherwise aggressive cancer.

All hCG diagnostic tests are carried out as immunoassays, based on specific antibodies that are able to distinguish hCG from three other, closely related, members of the glycoprotein hormone family (luteinizing hormone, LH; follicle-stimulating hormone, FSH; and thyroid stimulating hormone, TSH) [173]. The α-chains of these four hormones are identical, and are each comprised of 92 amino acids. The β-chains however are hormone specific and determine the particular activity of each hormone [174], but exhibit significant sequence homology between all four β-chains, with LH having 86%, TSH 46% and FSH 36% commonality with the first 114 amino acids of hCG [175]. Both the α- and β-chains are composed as three loops held in place by a "cysteine knot" of three disulphide bonds [4, 176]. Due to this distinctive molecular architecture, the glycoprotein hormones are classified within the cysteine knot growth factor (CKGF) family of proteins, even though they perform as hormones, rather than growth factors.

These similarities in sequence and structure, especially within the quartet of the glycoprotein protein hormones (hCG, LH, TSH, FSH), make stringent demands of antibody specificity when immunoassays are used to distinguish between the different molecules. To ensure high fidelity in hCG detection and quantification, rigorous antibody characterisation and selection is essential and, for this reason, much attention has been paid to hCG immunochemistry in terms of defining epitope structure and relating antibody binding specificity to structural differences in the various isoforms of hCG. The diagnostic value of any hCG immunoassay depends totally on the properties of the antibody around which the assay is built. Previous analysis of the hCG epitope structure [177], focused on the assignment of epitopes to the 3D structure of hCG, together with a clear linkage to identified monoclonal antibodies.

However, a wide range of hCG variants exist differing both in glycosylation state and fragment size. While hCG can exist as the normal α/β hetero-dimer, as individual subunits, or even as α/α homo-dimers. The most frequently encountered fragments are from the β subunit, in the form of nicked hCG β-subunit (hCGβn), hCG β-core fragment and hCGβ lacking the C-terminal peptide (-CTPhCGβ) [167, 178, 179]. This heterogeneity poses a major obstacle to the

development of a quantitative universal immunoassay for hCG [180], especially as many hCG epitopes depend on the molecule being intact and/or in its native conformation [175]. An assay architecture built around a single epitope site is preferred, as a two-site assay (double antibody sandwich) could result in more false negative results due to the doubled risk of one of the necessary epitopes being absent in a hCG variant [181]. However, the performance of a single site assay depends on the identity and nature of the chosen epitope, and it is crucial that it is present, intact and correctly folded in every hCG variant encountered.

It is against this background that the β3-loop of hCGβ_{66-80} (Fig. 5.1) has been identified as a potentially ideal epitope on which to base universal, single-site hCG assays. This loop is present in all known variants of hCG, and we postulate that it might have its own, self-contained secondary structural features that are retained, regardless of the conformation or fragmentation state of the rest of the molecule. It is suggested here that this local structure provides the basis for antibodies against this epitope to bind with similar affinity to any known hCG variant, as long as the relevant linear sequence remains accessible.

5.2 Experimental

5.2.1 Peptide Synthesis

Peptides were synthesised by *Mologic Ltd.* with *N*-(9-fluorenyl)methoxycarbonyl (Fmoc) - chemistry on TGA resin (Merck Chemicals) using a Liberty Microwave Synthesiser (CEM). Completed peptide resins were cleaved in 95% TFA (Rathburn), 2.5% water and 2.5% trisopropylsilane (Sigma) for two hours and recovered TFA liquors were dried by rotary evaporation and precipitated in cold tert-butyl methyl ether (Sigma) to afford a white solid. Freeze dried crude peptides were purified using a Dionex Ultimate 3000 HPLC system with an Onyx C18 column (Phenomenex) and a mobile phase gradient of 5% acetonitrile/water (0.1% TFA) to 100% acetonitrile (0.1% TFA) over 8 minutes. Pure peptide fractions were reduced by rotary evaporation and freeze dried from 50% acetonitrile. Purified peptides were analysed using a C18 kinetix column (phenomenex). Integrated peak areas typically indicated purity of 95% or above. Peptides were positively identified by Electrospray Mass Spectrometry (Waters ZMD MK II).

5.2.2 Screening assays

A range of peptide arrays were synthesised by *Mologic Ltd.* on polypropylene 455 well PEPSCAN cards using standard Fmoc-chemistry. Initially, to determine which linear sequential epitopes were recognised by various antibody preparations, a library of overlapping 12-mer peptides representing the entire sequence of hCGβ was created [182]; the first well was endowed with an array of identical immobilised peptides consisting of amino acids 1-12, the second well carried peptides consisting of amino acids 2-13, the third 3-14 etc. In conjunction with further libraries constructed with longer peptides (up to 20-mer), these libraries were used to determine which host species (mouse, rabbit, or sheep) were able to produce antibodies that recognised the linear sequence of the β3-loop. From these findings a panel of candidate monoclonal antibodies with binding specificity for the β3-region were selected. These monoclonal antibodies were prepared by hybridoma techniques in the laboratories of *Bioventix Ltd.*, Farnham UK. Subsequently, once the recognition of the correct sequential epitope (hCGβ_{66-80}) had been confirmed, a positional scanning library incorporating all single natural amino acid changes across the peptide sequence SIRLPGCPRGVNPVV (of hCGβ_{66-80}) was used. After deprotection with TFA and scavengers, the cards were washed extensively with an excess of H_2O and sonicated in disrupt-buffer containing 1% SDS / 0.1% β-mercaptoethanol in PBS (pH 7.2) at 70°C for 30 minutes, followed by sonication in H2O for another 45 minutes. The binding of antibody to each peptide was tested in a PEPSCAN-based ELISA. The polypropylene cards containing the covalently linked peptides were incubated with primary antibody 8G5 (obtained from sheep) diluted in blocking solution, 4% horse serum, 5% ovalbumin (w/v) in PBS / 1% Tween. After the wells had been washed, the peptides were incubated with a 1/1000 dilution of secondary antibody peroxidase conjugate for one hour at 25°C to detect and quantify the binding of 8G5 antibodies. After a further washing step, the peroxidase substrate 2,2'-azino-di-3-ethylbenzthiazoline sulfonate (ABTS) and 2μl of 3% H_2O_2 were added. Colour development was then measured by means of a charge coupled device (CCD) - camera and an image processing system, after a one-hour incubation [183].

5.2.3 Sample preparation

The peptides were dissolved directly in 10mM Phosphate buffer solution (pH 6.8), with a final concentration of 0.11 mM (0.177 mg ml^{-1}). The peptides stock concentration was confirmed by

UV-vis spectroscopy using a calculated extinction coefficient of ϵ_{214} = 21807 M^{-1} cm^{-1} [155]. For the samples with the fluorinated alcohol HFIP, water and 0.2 M buffer solutions were used to ensure the overall buffer concentration remained 10 mM.

5.2.4 Circular dichroism (CD) measurements

Far-UV CD spectra were recorded on a Jasco J-810 spectropolarimeter (Japan Spectroscopic Co., Tokyo) fitted with a Peltier unit for temperature control. The spectra are the average of three accumulations in step mode (data pitch of 1 nm, response time of 8s, bandwidth of 1 nm) acquired in rectangular cuvettes of 0.1 cm path length (Hellma GmbH, Germany) at 20°C. For every sample, an appropriate solvent spectrum was recorded under identical conditions and subtracted from the corresponding sample spectrum. The data collected was expressed in molar ellipticity (M^{-1} cm^{-1}).

5.2.5 Fourier Transform Infrared (FTIR) Spectroscopy

FTIR - attenuated total reflectance (FTIR-ATR) spectra were collected using a Tensor-37 FTIR spectrophotometer (Bruker Optics) equipped with a thermostated BioATR II unit and a liquid nitrogen cooled photovoltaic mercury cadmium tellurium (MCT) detector. The spectra were analysed using the OPUS software (Bruker Optics). Five acquisitions (128 scans, 4 cm^{-1} resolution) were done at different peptide concentrations. The contribution from atmospheric gasses was eliminated using an algorithm provided by the OPUS software. The contribution of liquid water was subtracted from the peptide spectrum using the combination band of water, centred at 2125 cm^{-1}, by flattening the region from 1906 to 1740 cm^{-1}.

FTIR spectra of synthetic peptides should be analysed carefully due to the possibility of trifluoroacetic acid (TFA) being present as an impurity in the sample, resulting from the RP-HPLC purification step. It is important to consider TFA, as it is known to absorb in the Amide I region at 1673-1680 cm^{-1} and can result in an overestimation of certain secondary structural elements (e.g. turns). In order to compensate for this, the TFA contribution is subtracted if the spectrum of a synthetic peptide presents a peak centred around 1200 cm^{-1} [184].

Peptide Sequence	Counter ions	Notes
SIRLPGCPRGVNPVV	2 Chloride	(Fig. 5.2b).
SIRLPGCPRGVDPVV	1 Chloride	
SIRLPGCPRGVHPVV	2 Chloride	
SIRLPGCPRGVLPVV	2 Chloride	
SIRLPGCPRGVGPVV	2 Chloride	

Table 5.1: Peptide sequences and counter ions used in unit cells.

5.2.6 Computational Method

An array of peptides was generated by performing point mutations on the structure of hCGβ_{66-80} obtained from 1HRP.pdb (Fig. 5.2), varying only the coordinates of the side chain and not the backbone. From these, unit cells were set up, which comprised a peptide sequence, 1853 water molecules and counter ions in order to ensure charge neutrality (Tab. 5.1). The molecules were arrayed on a cubic lattice with random molecular orientations in a cubic box of length 38.2Å. The rigid, non-polarisable, 3-site TIP3P empirical force field was used to describe the water molecules, the peptide was described by the CHARMM22 force field [31], while the PINY simulation package [99] was employed to perform the simulations.

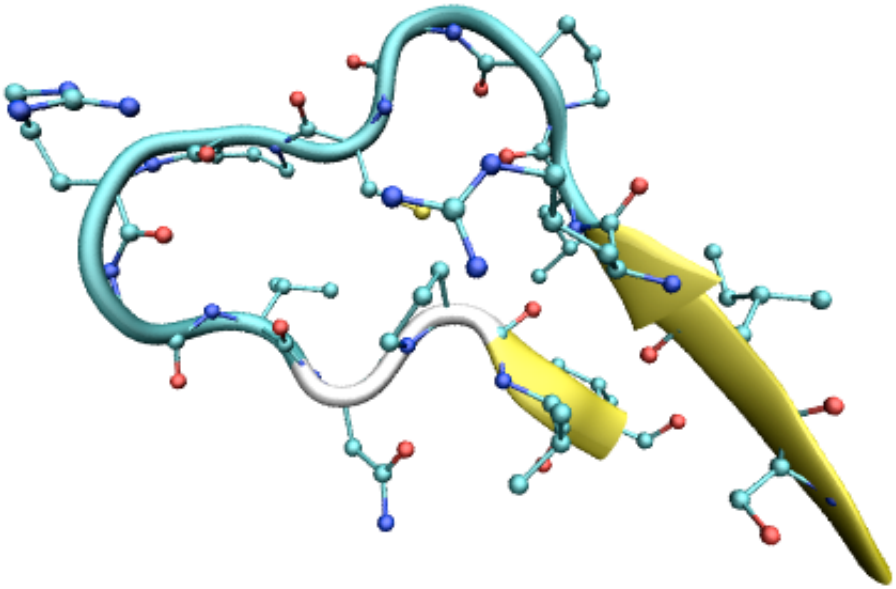


Figure 5.2: Snapshot of the initial structure of hCGβ_{66-80} obtained from 1HRP.pdb [4]. Note: Hydrogen atoms omitted.

The systems were equilibrated at 300K in the canonical ensemble for 350ps with a time-step of 0.5fs to anneal out unphysical contacts. The system was then run for 300ps with a time step of 0.5fs in the isothermal-isobaric ensemble to allow for spatial relaxation. The system was then run for a further 150ps in the canonical ensemble with a time-step of 1fs being employed, before a production run of 15ns using a time-step of 2.5fs from which all data was collected. Periodic boundary conditions were employed and long-range interactions were evaluated via Ewald summation.

5.3 Results

5.3.1 Mutational analysis

Of the three species tested, only sheep were found to produce antibodies to the β3-loop, with the loop appearing to constitute an immunodominant epitope. The sequence of the β3-loop is highly conserved with hCGβ_{66-80} and LHβ_{86-100} differing by only one amino acid; asparagine (N) in hCG *cf.* aspartate (D) in LH (Fig. 5.3a). Utilising a particular sheep monoclonal antibody designated 8G5, point mutations in peptide arrays were introduced in the sequence SIRLPGC-PRGVNPVV (hCGβ_{66-80}), to determine whether substitution of the asparagine (Asn_{77}) residue (Fig. 5.3b) had an impact on the antigen-antibody binding.

Mologic Ltd. found that this single amino acid difference between LH and hCG conferred a substantial specificity bias to the antigen-antibody interaction (Fig. 5.3), in favour of hCG with all five of the monoclonal antibodies binding to this region. Careful calibration of the LH cross-reactivity (with intact LHβ and hCGβ) revealed this to be 2.5% for 8G5 (additional antibodies were tested, and this cross-reactivity was found to range between about 0.27% and 2.5%). Substitution involving histidine (H) was the only other mutation found to retain significant binding.

In view of the potential usefulness of this particular epitope-antibody combination, we explored the structure of the β3 epitope recognised so strongly by the ovine immune system by means of a combination of experimental techniques and computer simulation in the free peptide. By this means we sought to determine if there are structural differences between these two peptides; such structural variations may be evidence of an inability of the LHβ version of the epitope to bind successfully with the paratope. Additional point mutations were investigated,

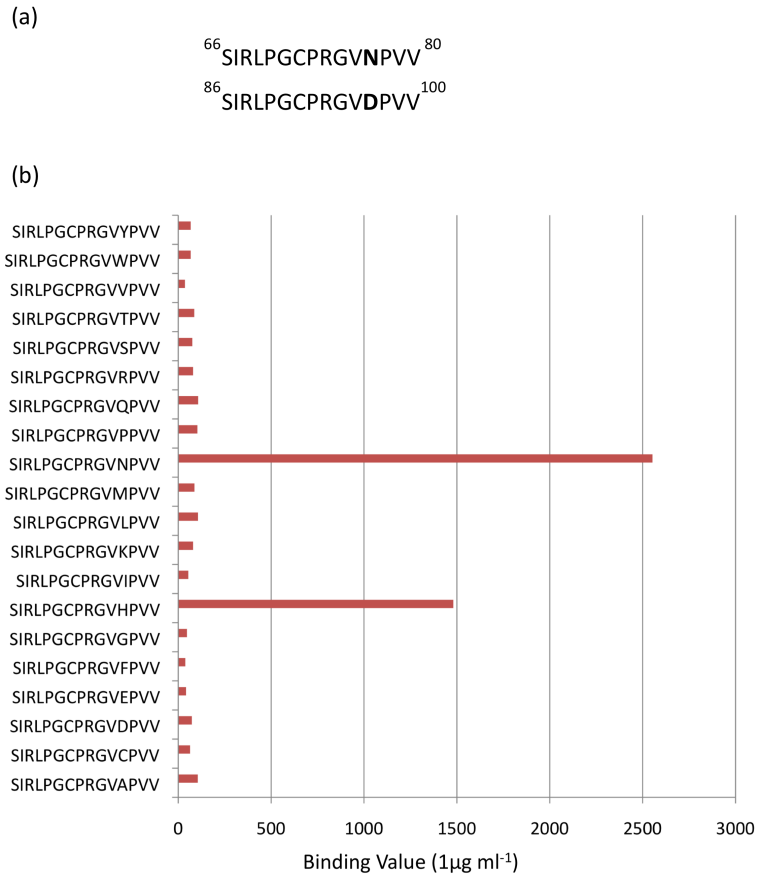


Figure 5.3: (a) Primary structures of hCGβ_{66-80} (upper) and LHβ_{86-100} (lower). (b) Binding results from a small subset of the Pepscan positional substitution library, in which there is a systematic substitution of position 77, occupied by asparagine in hCGβ. Note that there is negligible binding to the peptide occupying the position third from the bottom, which equates to the sequence in LH. In the sequence containing histidine at position 77, the binding is $\sim$60% that of the hCGβ binding.

involving substitution of a proline (Pro$_{78}$) residue [185]. The results revealed that binding to the monoclonal antibody 8G5 was lost.

5.3.2 Secondary structure of hCG

5.3.2.1 Structure predicted by simulations

In the crystal structure of hCGβ, hCGβ_{66-80} is defined by two turns with β-sheet structure over the N- and C-termini of the peptide (Fig. 5.2). A molecular dynamics simulation was performed to reveal the native structure of hCGβ_{66-80} under aqueous conditions (Fig. 5.4), in order to establish whether the structural features found in the crystalline sub-unit were retained in the

free peptide.

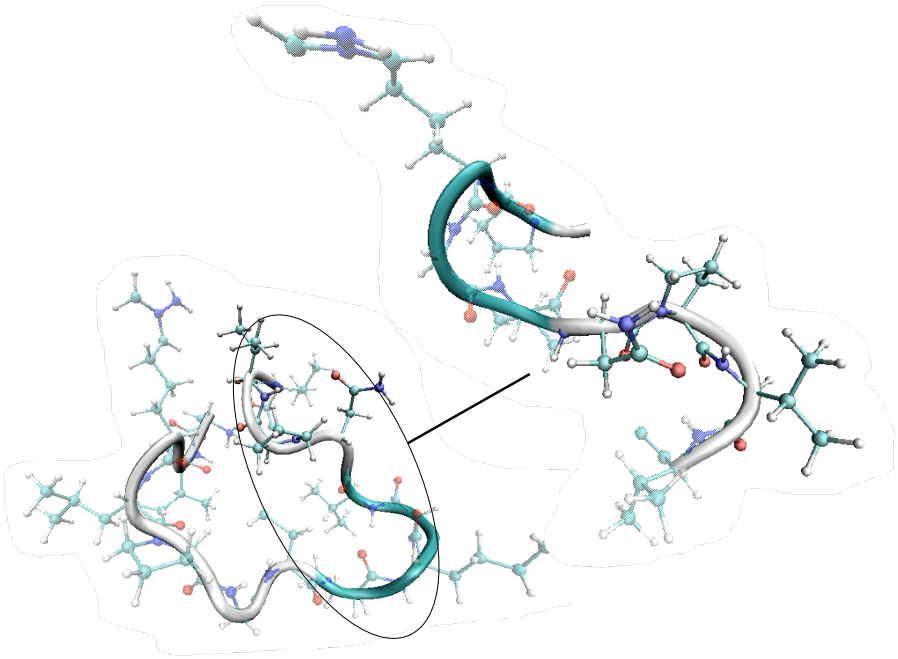


Figure 5.4: Secondary structure of hCGβ_{66-80} after 15ns.

The secondary structure of each residue as a function of time was analysed utilising the STRIDE secondary structure assignment algorithms [107] (Fig. 5.5). From this, it was clear that the simulation predicted a stable turn motif spanning the residues Pro_{73}-Val_{76}, as evident from it persisting over the course of the 15ns simulation. The analysis conducted utilising STRIDE also revealed that the turn structure spanning residues Leu_{69}-Cys_{72} was not as stable, with the sequence becoming more unordered by 5ns into the simulation; STRIDE did predict the turn structure to return sporadically. Furthermore, evidence of β-sheet character in the full protein (present as an initial condition in the free peptide between Arg_{68} and Val_{79}) was found to anneal away very quickly. This aspect of the full protein was therefore not retained in the isolated peptide.

A turn can be defined as the close approach (< 7Å) of two C^{α} and it does not necessarily have to have a well formed hydrogen bond. To that end, the separation of all C^{α} involved in γ-turns, β-turns, α-turns, and π-turns (separated by 2, 3, 4 and 5 peptide bonds respectively) were calculated over the last 5ns of the simulation. The results revealed that there was

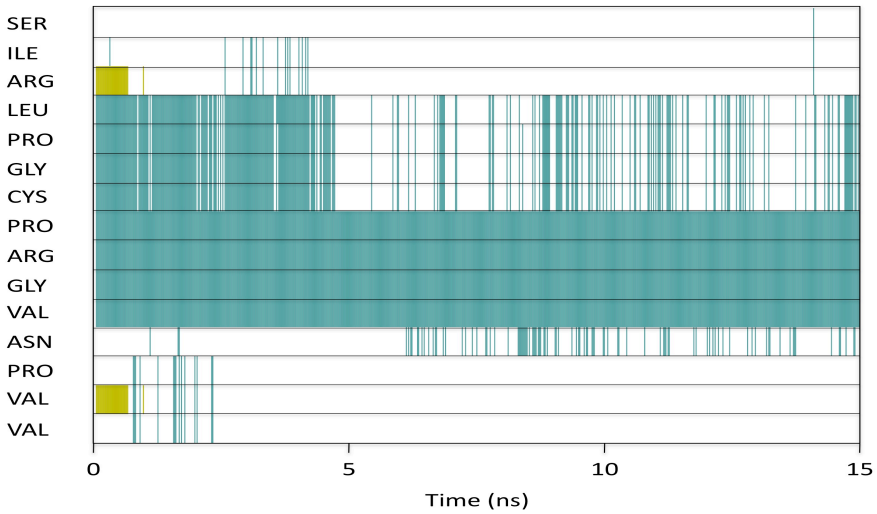


Figure 5.5: Secondary structure of each residue of hCGβ_{66-80} as a function of time. Green corresponds to turn, white is unordered and gold is β-strand. Simulation predicts a stable turn over Pro_{73}-Val_{76}, whereas a second turn spanning Leu_{69}-Cys_{72} was found to be only a transient feature.

a clear indication of a β-turn spanning residues Pro_{73}-Val_{76} (Fig. 5.6) as was predicted by the STRIDE algorithms, with the entirety of the C^{α} separation distribution falling below the assumed 7Å threshold. Furthermore, bond probability distribution analysis revealed strong evidence of a β-turn stabilising hydrogen bond between the backbone carbonyl group of Pro_{73} and the backbone amine group of Val_{76}, with the peak at a distance less than the assumed Hydrogen-bonding threshold of 2.5Å [106].

The potential presence of a β-turn motif spanning residues Leu_{69}-Cys_{72} was also partially confirmed by the separation of the C^{α} of Leu_{69} and Cys_{72} (Fig. 5.7a). While the peak was positioned above the assumed 7Å threshold, a notable portion of the distribution was found to lie below this cut-off. This appeared to support the findings of the STRIDE algorithm, by revealing that a β-turn structure spanning these residues may appear to some degree over the course of the last 5ns of the simulation; STRIDE revealed sporadic appearances of this motif over this same time period (Fig. 5.5). However analysis of the bond probability distribution involving the backbone carbonyl group of Leu_{69} and the backbone amine group of Cys_{72} revealed a peak at 3.66Å (Fig. 5.7a), well above the assumed 2.5Å cut-off. This therefore ruled out the presence of a stabilising hydrogen bond between these atoms, and may explain why the turn

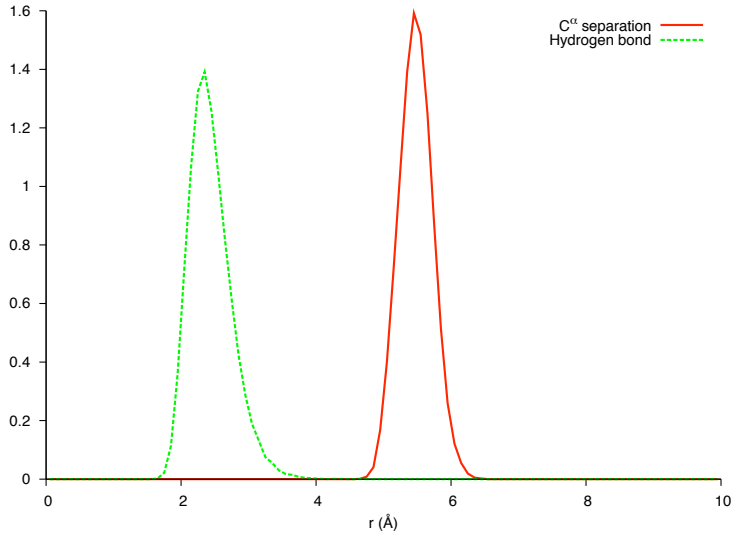


Figure 5.6: Solid (red line), separation of C^{α} in Pro_{73} and Val_{76}. Dashed (green line), separation between backbone carbonyl (Pro_{73}) and amine (Val_{76}) groups.

was only detected sporadically. Similarly, while the C^{α} belonging to residues Asn_{77} and Val_{80} were found to be separated at a distance below the 7Å cut-off (Fig. 5.7b), the separation between the backbone carbonyl and amine groups (3.86Å) was too great to support the presence of a stabilising hydrogen bond.

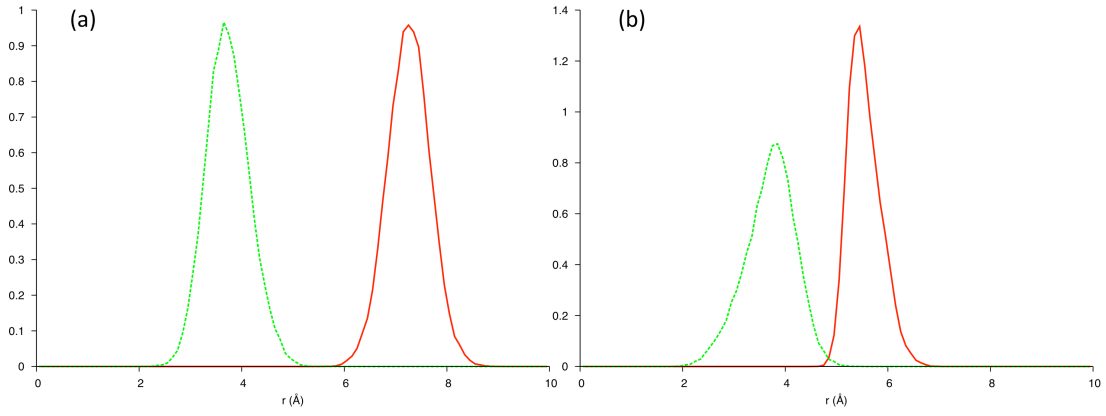


Figure 5.7: Solid (red line), separation of C^{α}. Dashed (green line), separation between backbone carbonyl and amine. (a) Leu_{69}-Cys_{72}, (b) Asn_{77}-Val_{80}.

Every pair of potential C^{α} involved in γ-turns presented a peak below the 7Å threshold (Fig. 5.8a) therefore it was necessary to evaluate the bond probability distributions of the backbone carbonyl group of residue i and the backbone amine group of residue $i+2$ (Fig. 5.8b). For every pair of residues evaluated, the separation of the two atoms was found to be greater than the assumed 2.5Å threshold therefore precluding the existence of stable γ-turns in the structure of hCG β_{66-80}. Analysis of the separation of C^{α} involved in α-turns (Fig. 5.9a) resulted in only one pairing (Cys_{72}-Val_{76}) presenting a peak below the assumed 7Å cut-off (6.27Å), however bond probability analysis of potential hydrogen bonds ruled out the existence of a stable α-turn between these residues due to the corresponding peak falling at a separation of 5.36Å (Fig. 5.9b).

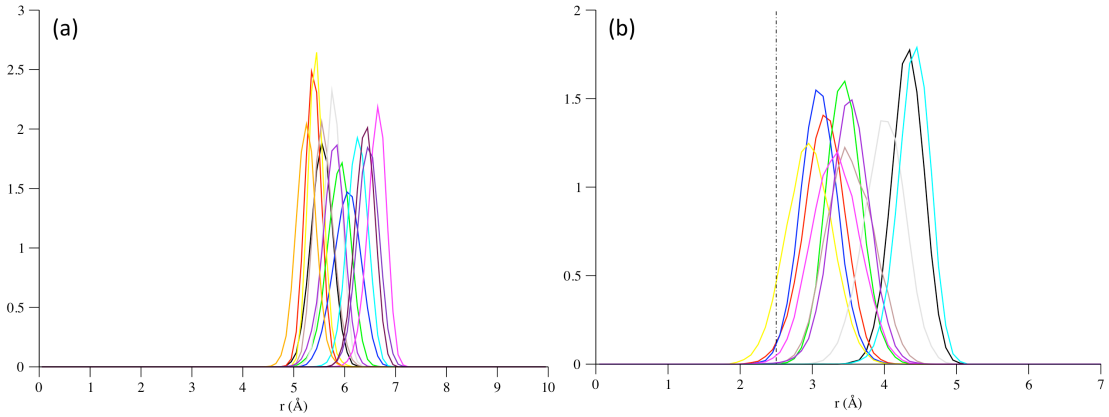


Figure 5.8: (a) Separation of C^{α} involved in γ-turns in hCGβ_{66-80}, (b) Bond probability distribution analysis for potential hydrogen bonds.

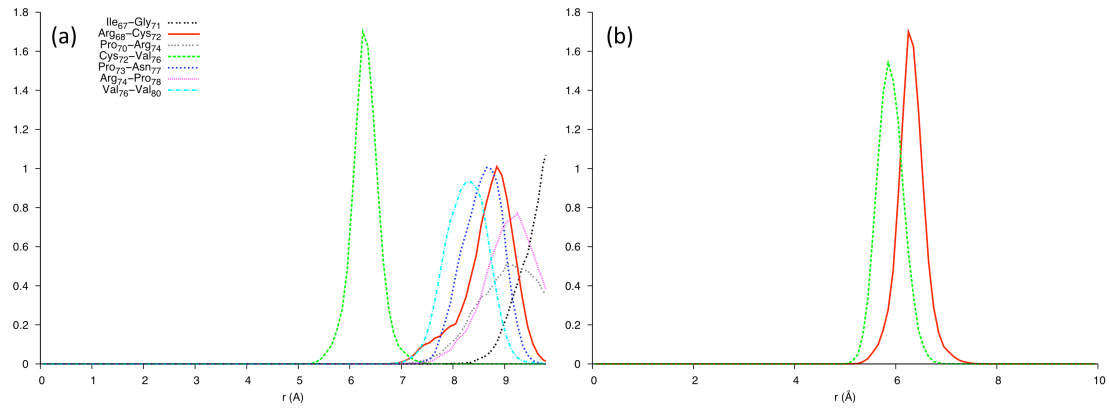


Figure 5.9: (a) Separation of C^{α} involved in α-turns in hCGβ_{66-80}, (b) Solid (red line), separation of C^{α} in Cys_{72} and Val_{76}. Dashed (green line), separation between backbone carbonyl (Cys_{72}) and amine (Val_{76}) groups.

Finally analysis of the separation of C^{α} involved in π-turns resulted in negative results for all possible permutations with no evidence of such motifs existing in the secondary structure of hCG β_{66-80} (Fig. 5.10).

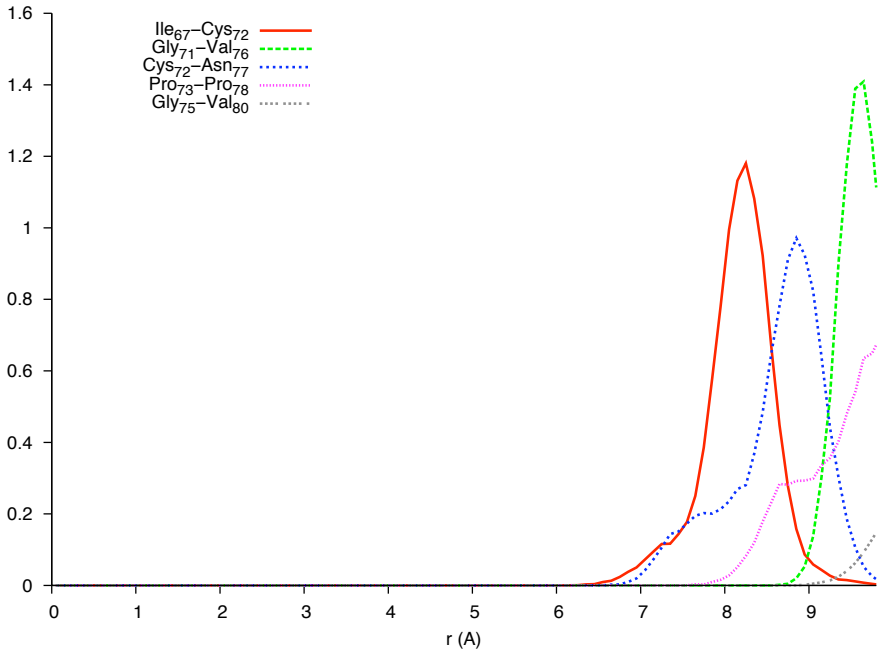


Figure 5.10: (a) Separation of C^{α} involved in π-turns in hCGβ_{66-80}.

The simulation therefore led to a model of the free peptide in which a single stable turn coexisted with other transient structures; more open and less structured that the sequence in the full β-chain. Direct experimental confirmation of such a dynamic situation would be challenging, yet it was important to establish whether the simulations captured the general secondary structure of the free peptide.

5.3.2.2 Spectroscopic analysis

The measured FTIR spectra of hCGβ_{66-80} was found to exhibit a very broad Amide I band, indicative of the co-existence of different conformations in solution (Fig. 5.11). Over the replica measurements, the second derivative spectra revealed bands consistent with the presence of a fractional population of turns in solution [63]. The position of the Amide II band (maximum at 1551 cm^{-1}) also suggested that β-turns were present as the band was positioned at a higher frequency when compared to other structural elements [186].

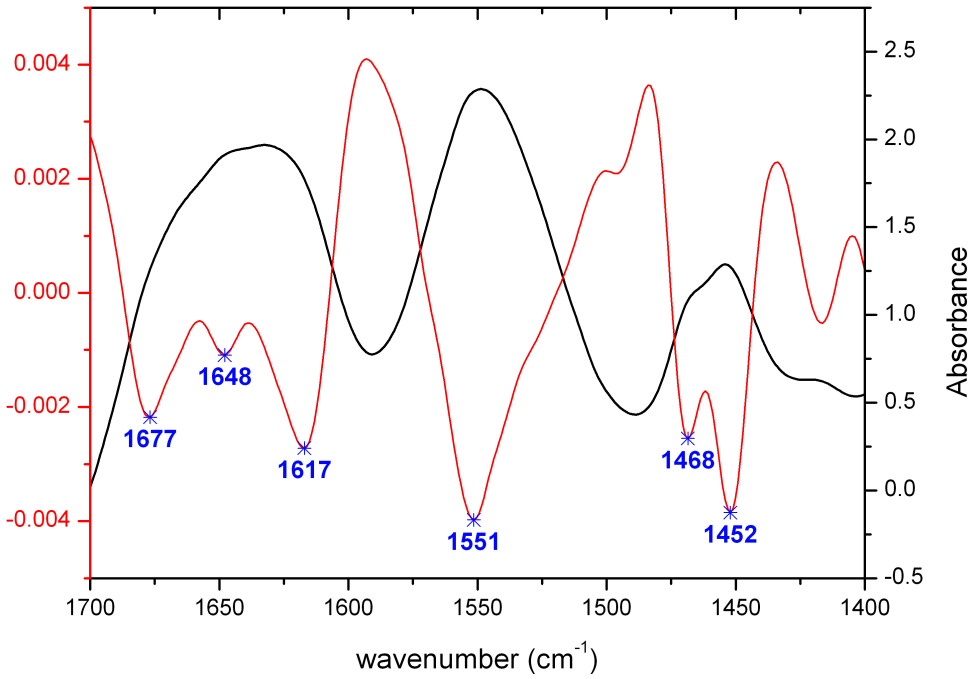


Figure 5.11: FTIR-attenuated total reflectance blank-subtracted and TFA-corrected spectrum of hCGβ_{66-80} (black) displaying the Amide I (maximum at 1633 cm^{-1}) and Amide II (maximum at 1551 cm^{-1}) regions, and the second derivative spectrum (red) displaying the peak positions of the different components.

Evidence for turns in CD data is more ambiguous; a turn can give variable signatures in CD with bands of positive or negative ellipticity at wavelengths of approximately 200nm. In particular, type II β-turns should exhibit a weak CD band with negative ellipticity at 220-230nm, with a stronger positive band between 200-210nm. Type I β-turns on the other hand give CD spectra which are qualitatively similar to that of an α-helix; characterised by a band of positive ellipticity near 190nm and two bands of negative ellipticity at longer wavelengths [45]. In order to confirm if a turn structure was indeed present in the secondary structure of this peptide, far-UV CD spectra of hCGβ_{66-80} were recorded (Fig. 5.12). As evident from the recorded CD spectra, none of the typical signatures of a turn structure were present, and the single CD band with negative ellipticity near 200nm was more indicative of a disordered structure. This suggested that the turn was not a stable secondary structure element for this peptide in an aqueous buffer, or if the turn was present it was part of an ensemble of open conformations.

To examine whether the conformational bias of the ensemble could be altered to enhance the population of a particular fold type, experiments were performed in fluorinated alcohol HFIP-

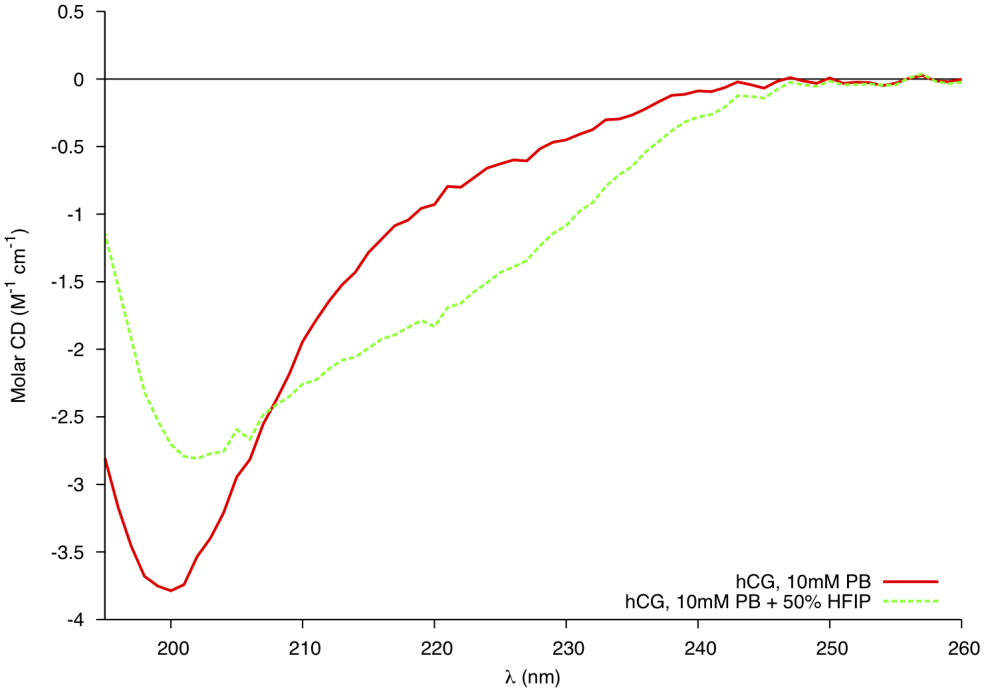


Figure 5.12: Far-UV CD spectrum of 0.11mM hCG β_{66-80}.

water mixtures. Fluorinated co-solvents exclude water from the peptide's environment, favouring the formation of intramolecular hydrogen bonds. Therefore, if a hydrogen bond was responsible for the stabilisation of a particular conformation, a change in solvent conditions would cause a shift in the fractional population of conformers which would be observed in a CD spectra due to the quantitative nature of CD. Far-UV CD spectra of hCGβ_{66-80} in 50% HFIP (v/v) were recorded (Fig. 5.12). The presence of the Pro residues in the sequence of hCGβ_{66-80} would be expected to strongly disfavour the formation of a helical motif (Tab. A.4). Coupled with the structural preferences predicted by MD, this observed change in the spectrum could be indicative of the stabilisation of a turn motif.

On the basis of these spectra, it was concluded that the secondary structure of hCGβ_{66-80} was comprised of an ensemble of open structures and turn motifs and that the simulated configurations (Fig. 5.4) were consistent with the experimental measurements.

5.3.3 Secondary structure comparison between LH and hCG

5.3.3.1 Spectroscopic analysis

Experimentally, the FTIR spectra of LHβ_{86-100} (Fig. 5.13) was found to be similar to that observed in hCGβ_{66-80} (Fig. 5.11). Namely a broad Amide I band was found which indicated the co-existence of different conformations in solution, and the second derivative spectra was consistent with fractional populations of turns [63]. The position of the Amide II band (maximum at 1553 cm^{-1}) also suggested that β-turns were present [186].

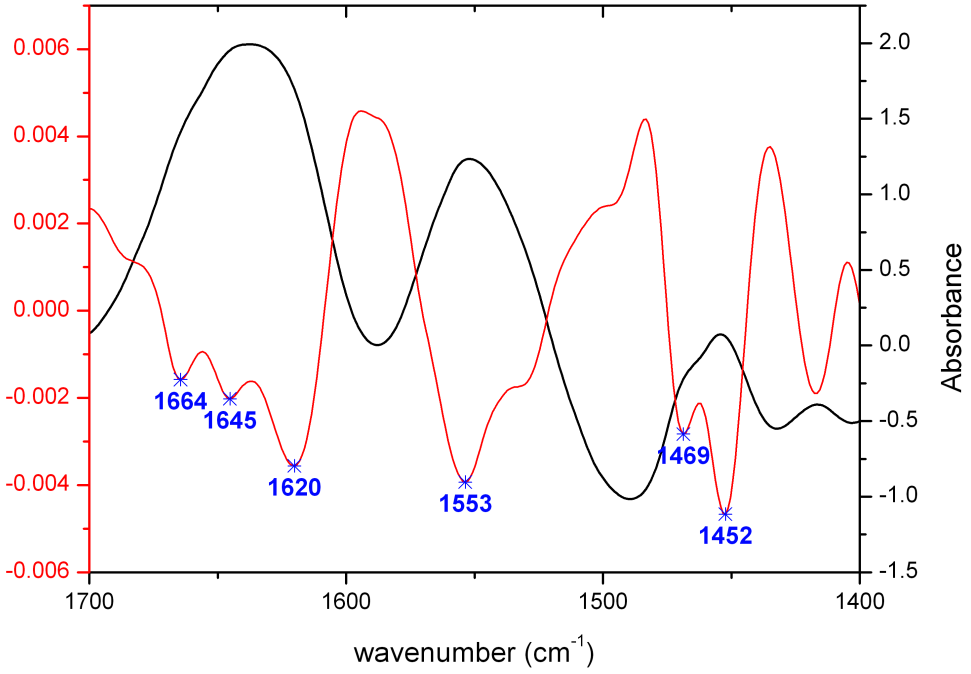


Figure 5.13: FTIR-attenuated total reflectance blank-subtracted and TFA-corrected spectrum of LHβ_{86-100} (black), and the second derivative spectrum (red) displaying the peak positions of the different components.

The recorded far-UV CD spectra of LHβ_{86-100} (Fig. 5.14) also revealed a predominantly disordered structure. There were however detectable differences between the hCGβ_{66-80} and LHβ_{86-100} spectra ($\sim$220nm) which implied that the single point mutation had influenced the conformational equilibrium. Far-UV CD spectra of LHβ_{86-100} in 50% HFIP (v/v) also suggested the stabilisation of a turn motif (Fig. 5.14), although potentially to a lesser degree than observed for hCGβ_{66-80}. The weaker shoulder of negative ellipticity at 220-230nm may be due to LHβ_{86-100} presenting weaker intensity than hCGβ_{66-80} at these wavelengths in aqueous buf-

fer. This may be indicative of hCGβ_{66-80} being comprised of more stable turn motifs than the LHβ_{86-100} mutant under such conditions.

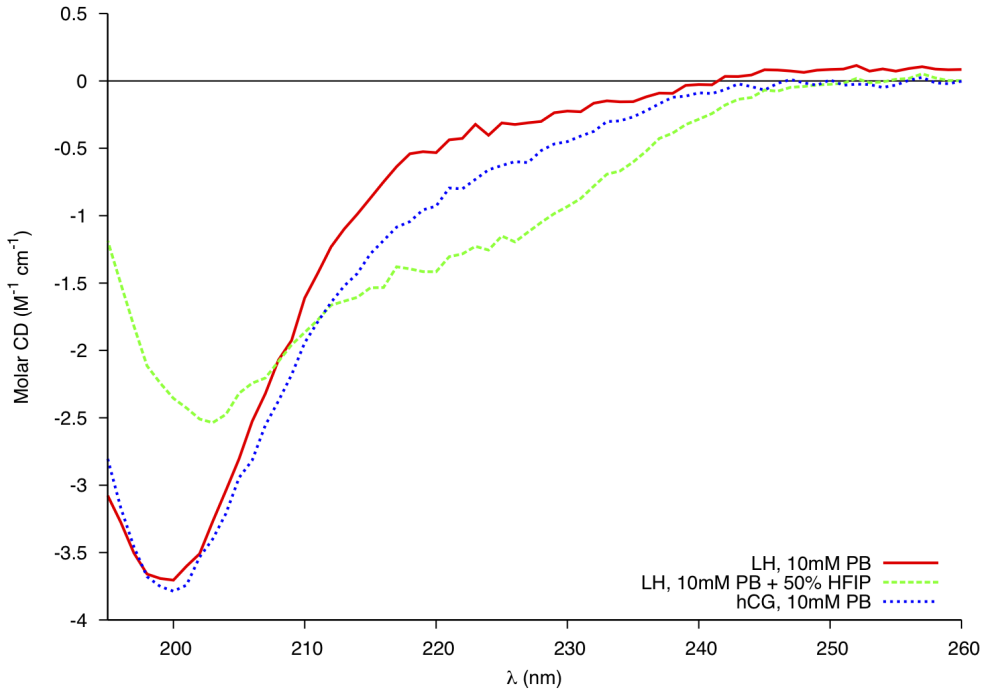


Figure 5.14: Far-UV CD spectrum of 0.11mM LHβ_{86-100}.

Unsurprisingly, the main structural features of the LHβ_{86-100} and hCGβ_{66-80} peptides were found to be similar, with each peptide comprised of an ensemble of open structures and turn motifs. The experimental data, however, also revealed evidence of subtle differences arising from the point mutation. Unfortunately, precise secondary structure interpretation was beyond the resolution of the techniques employed. As the MD simulations and experiments of hCGβ_{66-80} were within the resolution limits of the measurements, further emphasis was placed on a secondary structure comparison based on the simulation evidence.

5.3.3.2 Simulation

Molecular dynamics simulations were performed on the sequence emulating that found in the LHβ3 sub-unit, in order to investigate whether the mAb 8G5 molecular recognition mechanism and selectivity between hCG and LH was due to a structural variation induced by the single amino acid mutation. As with the simulation involving the hCG molecule, the secondary struc-

ture of each residue as a function of time was analysed utilising the STRIDE secondary structure assignment algorithms [107] (Fig. 5.15). The obvious major difference between the secondary structure of hCGβ_{66-80} and LHβ_{88-100} was the increase in the span of the turn motif, which appeared to have extended to include residues Asp$_{97}$ and Pro$_{98}$. Furthermore, evidence of β-sheet character between residues Arg$_{88}$ and Val$_{99}$ was found to anneal away more quickly than in hCG (Arg$_{68}$ and Val$_{79}$). However, it did return sporadically during the initial 3ns of the simulation.

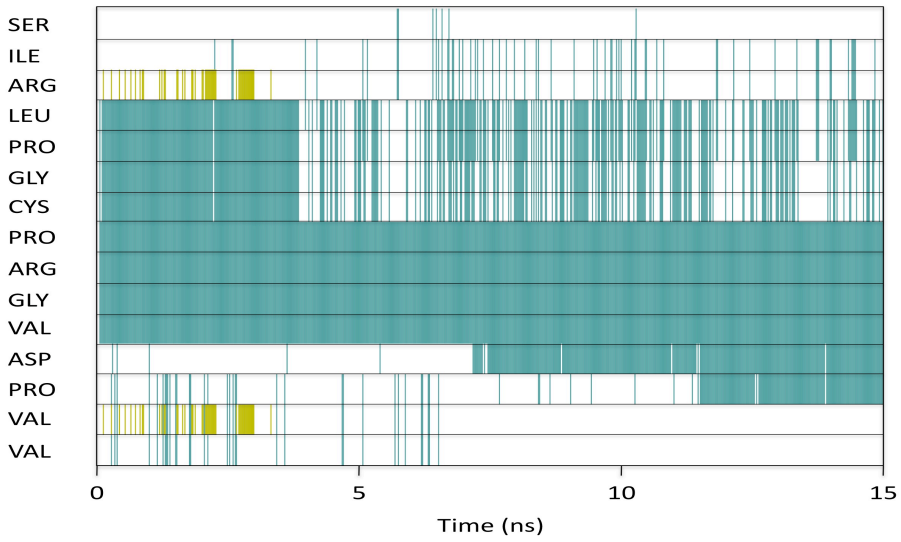


Figure 5.15: Secondary structure of each residue of LHβ_{86-100} as a function of time. Green corresponds to turn, white is unordered and gold is β-strand. Simulation predicts a stable turn over Pro$_{93}$-Val$_{96}$, with the span extending to include Asp$_{97}$ and Pro$_{98}$. A second turn spanning Leu$_{89}$-Cys$_{92}$ was found to be less stable.

The separation of all C$^{\alpha}$ involved in γ-turns, β-turns, α-turns, and π-turns (separated by 2, 3, 4 and 5 peptide bonds respectively) were calculated over the last 5ns of the simulation. As with the hCG simulation, there was a clear indication of a β-turn spanning residues Pro$_{93}$-Val$_{96}$ (Fig. 5.16), with the peak of the C$^{\alpha}$ separation distribution being positioned at 5.365Å. This was supported by bond probability distribution analysis which revealed a β-turn stabilising hydrogen bond between the backbone carbonyl group of Pro$_{93}$ and the backbone amine group of Val$_{96}$, as evident through the peak positioned at 2.146Å.

Additional C$^{\alpha}$ pairs which presented a peak below the assumed 7Å threshold were those belonging to Leu$_{89}$-Cys$_{92}$, Arg$_{94}$-Asp$_{97}$, Gly$_{95}$-Pro$_{98}$, and Asp$_{97}$-Val$_{100}$ (Fig. 5.17a). Of these,

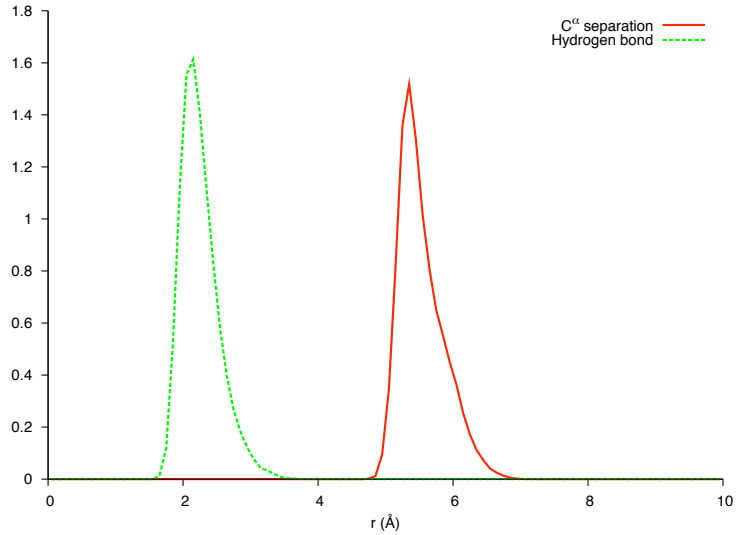


Figure 5.16: Solid (red line), separation of C^{α} in Pro_{93} and Val_{96}. Dashed (green line), separation between backbone carbonyl (Pro_{93}) and amine (Val_{96}) groups.

bond probability analysis revealed that only the Asp_{97}-Val_{100} pair coincided with a stabilising hydrogen bond between the backbone carbonyl and amine groups (Fig. 5.17b).

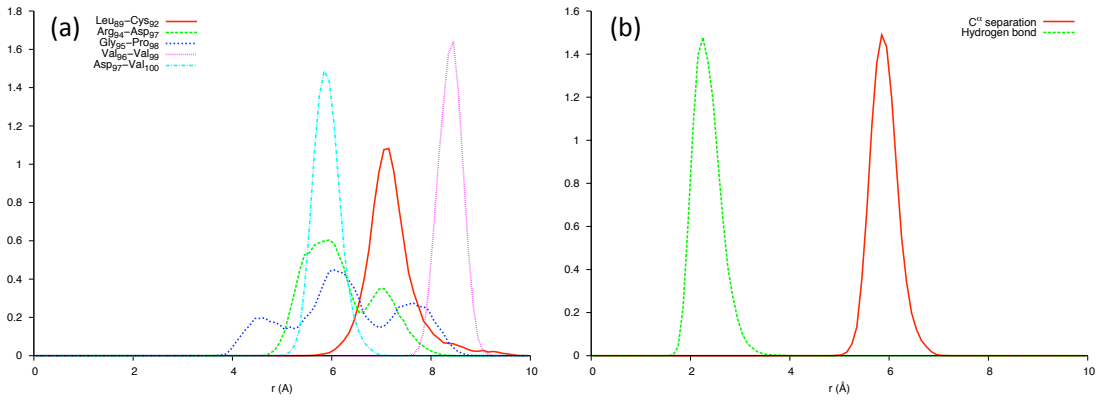


Figure 5.17: (a) Separation of C^{α} involved in β-turns (b) Solid (red line), separation of C^{α} in Asp_{97} and Val_{100}. Dashed (green line), separation between backbone carbonyl (Asp_{97}) and amine (Val_{100}) groups.

Visualisation of the secondary structure (Fig. 5.18) revealed that the sequence Pro_{93}-Val_{100} appeared to be comprised of sequential β-turns forming a 'S-bend' like motif. The peaks falling below 7Å for the C^{α} pairs of Arg_{94}-Asp_{97} and Gly_{95}-Pro_{98} must therefore be due to these

sequences forming part of this structure, rather than forming their own β-turn; interestingly, the C^{α} belonging to Val_{96} and Val_{99} were found to be separated by a distance greater than 8Å.

Figure 5.18: Visualisation of 'S-bend' like motif over Pro_{93}-Val_{100} in LHβ_{86-100}.

As in the structure of hCG β_{66-80}, every pair of C^{α} potentially forming γ-turns were found to have a separation less than the 7Å threshold and it was necessary to evaluate the bond probability distributions of the backbone carbonyl group of residue i and the backbone amine group of residue $i + 2$ (Fig. 5.19). And as in the hCGβ_{66-80} model, every pair of residues presented a negative result with the separation of the two atoms being found to be greater than the assumed 2.5Å threshold; ruling out the existence of stable γ-turns in the structure of LHβ_{86-100}.

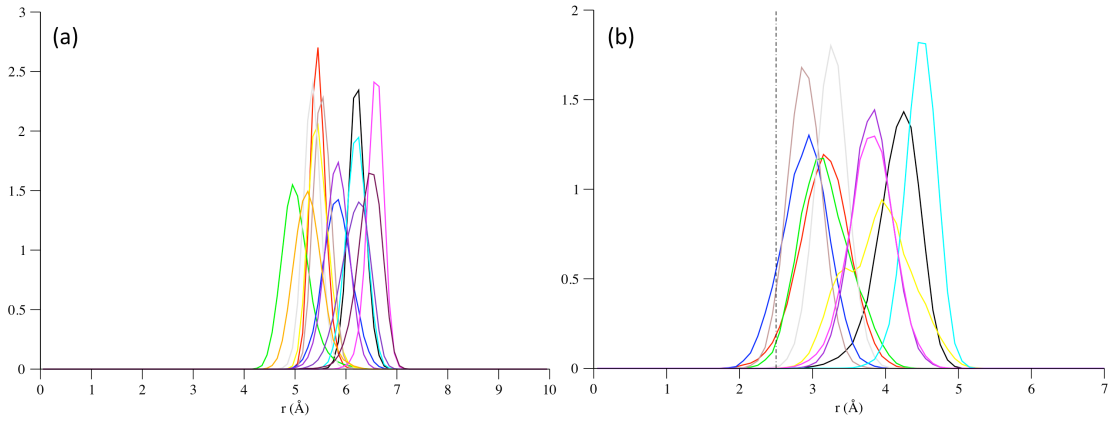


Figure 5.19: (a) Separation of C^{α} involved in γ-turns in LHβ_{86-100}, (b) Bond probability distribution analysis for potential hydrogen bonds.

The only pairs of C^{α} involved in potential α-turns to present peaks below the assumed 7Å threshold were those found in Cys_{92}-Val_{96} and Pro_{93}-Asp_{97} (Fig. 5.20a), however only the Pro_{93}-Asp_{97} pairing corresponded to a potential stabilising hydrogen bond (Fig. 5.20b). However due to the strong evidence of two β-turns spanning this sequence, the presence of an α-turn was unlikely. Finally, pairs of C^{α} involved in potential π-turns were evaluated. The pairs, Cys_{92}-Asp_{97} and Gly_{95}-Val_{100} were found to present peaks below 7Å, while Ile_{87}-Cys_{92} presented a peak with considerable intensity under this threshold (Fig. 5.21a). Bond probability analysis however revealed that none of the pairings coincided with evidence of stabilising hydrogen bonds (Fig. 5.21b), suggesting that such structures were in reality not present in the structure.

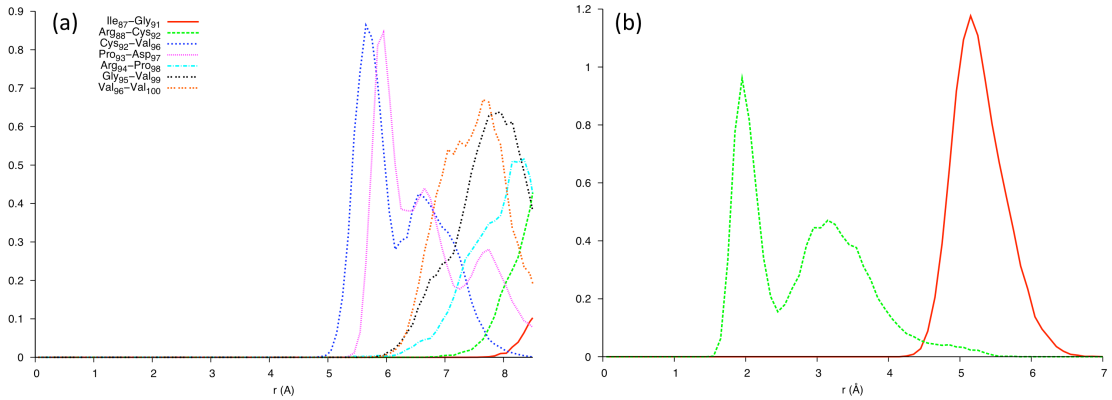


Figure 5.20: (a) Separation of C^{α} involved in α-turns in LHβ_{86-100}, (b) Bond probability distribution analysis for potential hydrogen bonds.

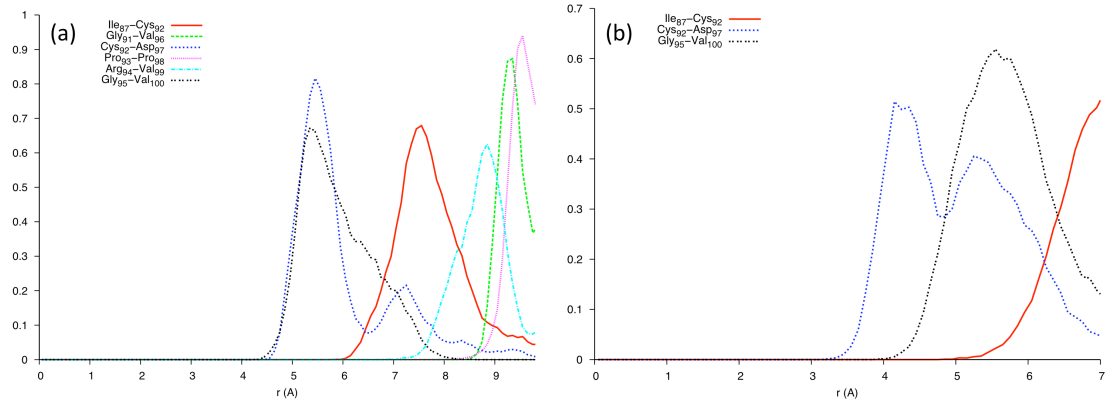


Figure 5.21: (a) Separation of C^{α} involved in π-turns in LHβ_{86-100}, (b) Bond probability distribution analysis for potential hydrogen bonds.

As the β-turn structure spanning residues Pro-Arg-Gly-Val appeared to be a common fea-

ture shared between hCGβ_{66-80} (Fig. 5.22a) and LHβ_{86-100} (Fig. 5.22b), it was important to determine if both turns were of the same type.

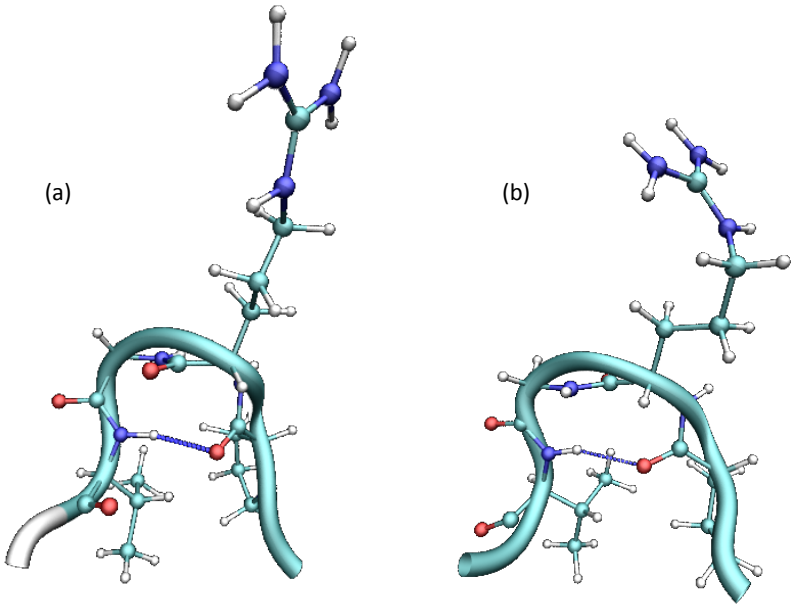


Figure 5.22: Visualisation of the turn motif spanning residues PRGV in (a) hCG, Pro_{73}-Val_{76}, and (b) LH, Pro_{93}-Val_{96}, revealing the stabilising hydrogen-bond between the backbone carbonyl group of the Pro residue and the backbone amine group of the Val residue

5.3.3.3 Finer details - turn

A turn can be classified by its backbone dihedral angles (Table. A.3); in the case of a β-turn, one focuses on the dihedral angles of the two central amino acids. The Ramachandran angles of the amino acids belonging to the sequence Pro_{73}-Val_{76} in hCGβ_{66-80} were calculated over the last 5ns of the simulation (Fig. 5.23). The two central amino acids were found to have Ramachandran angles (ϕ, ψ) equalling approximately -75°, -30° (Arg) and -85°, -10° (Gly). Based on these angles, the closest structure would therefore be the type I β-turn which is defined by the two central amino acids presenting Ramachandran angles of -60°, -30° and -90°, 0° respectively.

Over the last 5ns of the LHβ_{86-100} simulation, the torsional angles of Pro_{93}-Val_{96} (Fig. 5.24), however, revealed that the two central amino acids presented unique Ramachandran angles equalling approximately -60°, -120° (Arg) and 90°, 0° (Gly). This indicated a type II β-turn (known to present Ramachandran angles of -60°, 120° and 80°, 0° respectively) as opposed to the type I turn found for hCGβ_{66-80}.

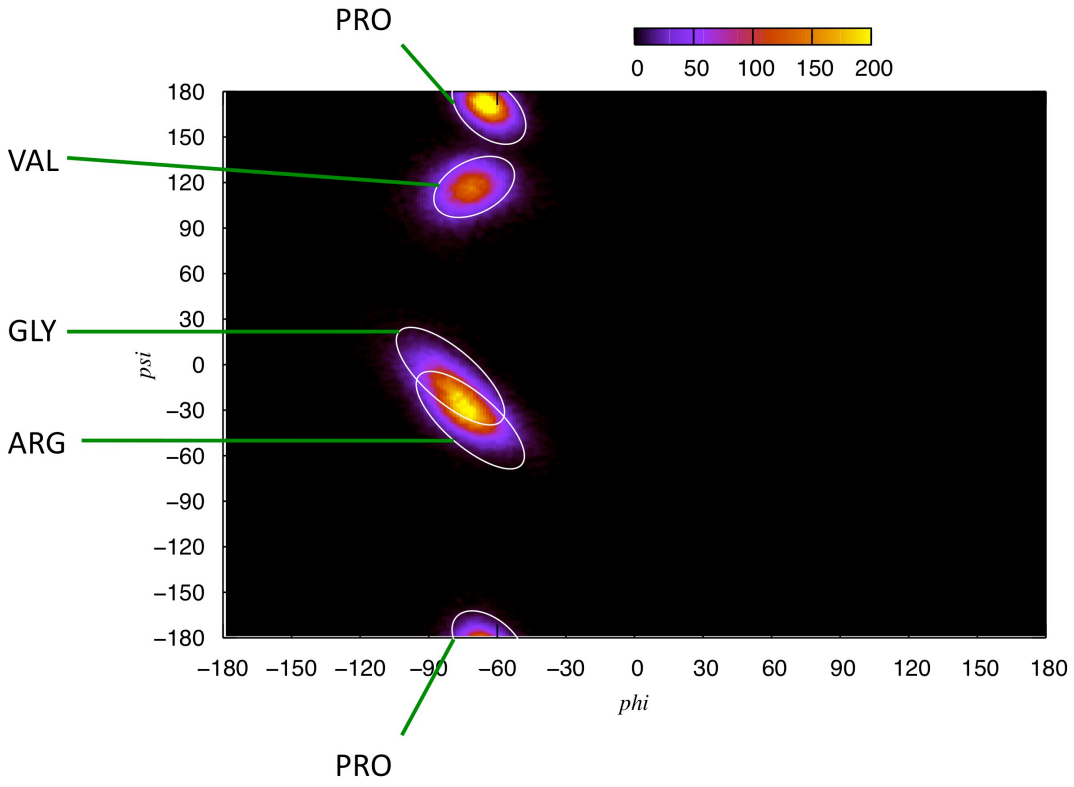


Figure 5.23: Ramachandran angles of sequence PRGV in hCGβ_{66-80} over the last 5ns. The torsional angles of the two central amino acids, Arg_{74} and Gly_{75}, indicate a type I β-turn.

The amino acids which comprise the proposed second β-turn (i.e. Asp_{97}-Val_{100}) however were found to present Ramachandran angles which indicated a type I β-turn (data not shown).

5.3.3.4 Finer details - clasp

Visualisation of the crystallographic structure of the hCGβ sub-unit (Fig. 5.25), reveals a hydrogen bond between the carbonyl group of the Asn_{77} side-chain and the backbone amine group of Val_{79} (Fig. 5.26a). It is known that asparagine can form hydrogen bonds with the polypeptide back bone, for example at the start or end of an α-helix or in turn motifs in β-sheets, and such a hydrogen bond could stabilise a hairpin loop whereby the peptide is “zipped" together at each end.

The accommodation of a substitution involving histidine can be rationalised by a similar orientation also allowing for this hydrogen bond (Fig. 5.26b). It is also believed that although

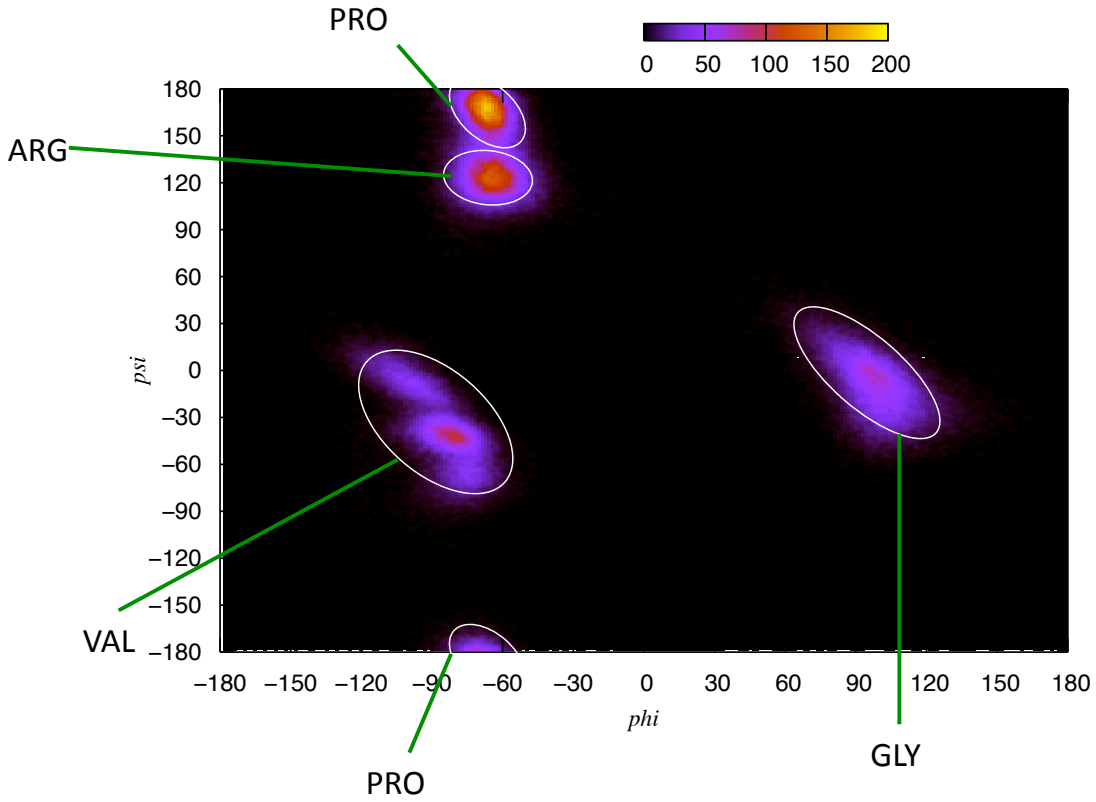


Figure 5.24: Ramachandran angles of sequence PRGV in LHβ_{86-100} over the last 5ns. The torsional angles of the two central amino acids, Arg_{94} and Gly_{95}, indicate a type II β-turn.

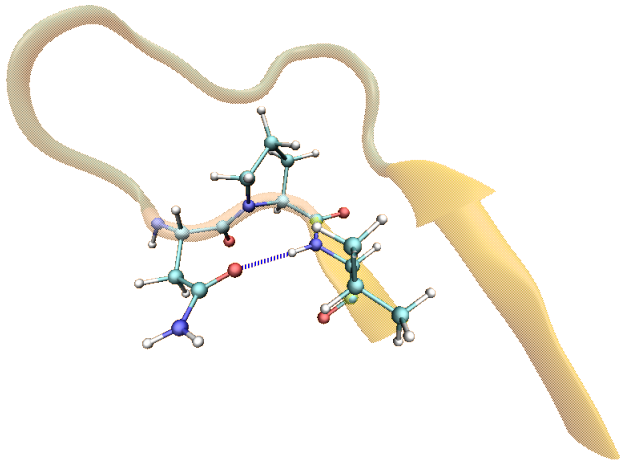


Figure 5.25: Crystallographic structure of hCGβ_{66-80} as obtained from 1HRP.pdb [4], revealing the hydrogen bond clasp.

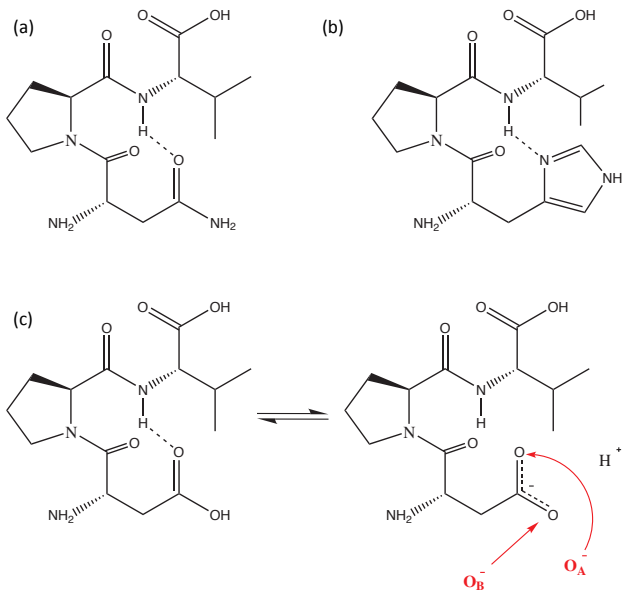


Figure 5.26: Proposed H-bonding geometry in (a) hCGβ_{66-80}, (b) His$_{77}$ mutant. (c) Asp$_{77}$ mutant depicting the proposed “clasp" being only present in the protonated form of Asp.

aspartate (native on LH) is very similar to asparagine (native in hCG), such a stabilising hydrogen bond would only be present if Asp were protonated (Fig. 5.26c). This seemed surprising, as one would have expected the deprotonated carboxylate anion to be a prime hydrogen bond acceptor. However due to the deprotonation, the oxygens and carbon of the carboxylate anion would be partially sp and sp^2 hybridised, respectively. This would lead to the formation of two partially double bonded C-O bonds, which could be spatially locked in an orientation disfavouring such an interaction.

In order to determine if such a hydrogen-bond was potentially present in the free peptide hCGβ_{66-80} during the course of the simulation, the bond probability distribution for these atoms was evaluated over the entirety and, separately, the last 5ns of the simulation (Fig. 5.27). Analysis of these distributions revealed that irrespective of the chosen time frame, two peaks were present and located at 2.158Å and 5.852Å. This indicated that while such a hydrogen-bond was present during the course of the simulation due to the presence of the peak below the 2.5Å threshold, such an arrangement was not stable over the entirety of the run and for considerable periods of times the separation of the atoms was too great to accommodate this proposed hydrogen bond. Overall the structure of this particular fragment was found to be much more dynamic than previously hypothesised.

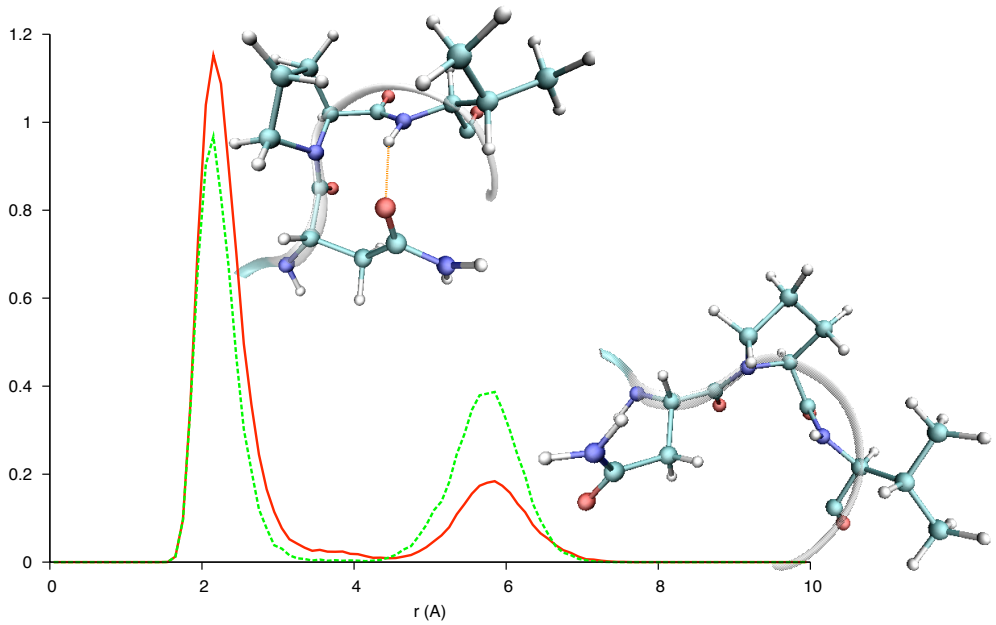


Figure 5.27: Bond probability analysis of the safety catch in hCGβ_{66-80}, coupled with visualisations of the structure corresponding to each peak. Solid (red) line: over full 15ns, Dashed (green) line: over last 5ns.

From the simulation of LH β_{86-100}, bond probability analysis was performed on each of the two carboxylate oxygens of the Asp_{97} side-chain (Fig. 5.26c); henceforth referred to as O_A^- and O_B^-; and the backbone amine group of Val_{99}. The bond probability analysis over the entirety of the simulation reveal a peak positioned at 5.662Å for the pairing involving O_A^- (Fig. 5.28a) and peak at 5.365Å, with a shoulder at 3.357Å for the O_B^- pair (Fig. 5.28b). Bond probability analysis over the last 5ns of the simulation reveals that these separations fluctuate to a greater degree, with the O_A^- pairing presenting a double peak at 5.662Å and 4.367Å (Fig. 5.28a), while the O_B^- pairing results in a peak at 5.365Å accompanied by two (more pronounced) shoulders at 4.319Å and 3.405Å (Fig. 5.28b).

Regardless of whether one focused on the last 5ns of the simulation or the simulation as a whole, neither carboxylate oxygen was found to present a peak below the assumed 2.5Å threshold. This therefore appeared to support the hypothesis that the stabilising hydrogen bond was not favoured in the presence of the deprotonated Asp_{97} residue. The lack of hydrogen bond clasp, coupled with the difference in turn type spanning the residues Pro-Arg-Gly-Val, would suggest that these structural features were important in the molecular recognition mechanism of the mAb

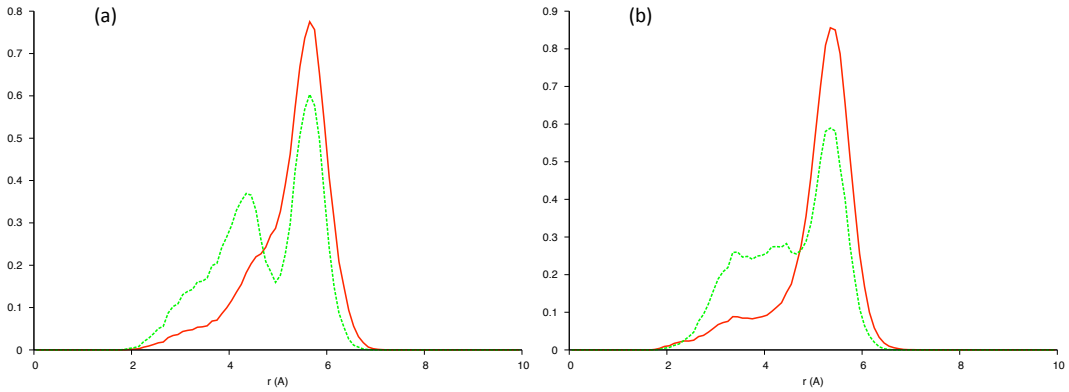


Figure 5.28: Separation of Asp_{97} carboxylate oxygen (a) O_A^-, (b) O_B^-, and hydrogen of the backbone amine group of Val_{99}. Solid (red) line: over full 15ns, Dashed (green) line: over last 5ns.

8G5 with the different β sub-units of hCG and LH.

5.3.4 Other mutations

In order to investigate whether these fine structural differences were indeed important in rationalising the 8G5 discrimination between hCG and LH, it was important to determine if these features were shared by other peptide systems. From the results of the point mutations (Fig. 5.3b), only the substitution involving histidine was found to retain significant binding to the monoclonal antibody and it was believed that such a substitution would result in an orientation accepting of a hydrogen bond. It was therefore of interest to probe if such a structural feature was indeed present in the structure of this peptide (SIRLPGCPRGV**H**PVV) and if the type I β-turn motif was a shared feature between this system and hCGβ_{66-80} - potentially explaining the retained binding to the mAb 8G5. To that end, a molecular dynamics simulation was performed on the sequence SIRLPGCPRGVHPVV.

As with the previous simulations, the secondary structure of each residue as a function of time was analysed utilising the STRIDE secondary structure assignment algorithms [107] (Fig. 5.29). The obvious major differences between the secondary structures was in the stability of the turn motif spanning residues Leu-Pro-Gly-Cys and in the β-sheet character between residues Arg and Val, with both structural motifs being more stable over the course of the 15ns simulation.

The Ramachandran angles of the amino acids belonging to the sequence Pro-Arg-Gly-Val

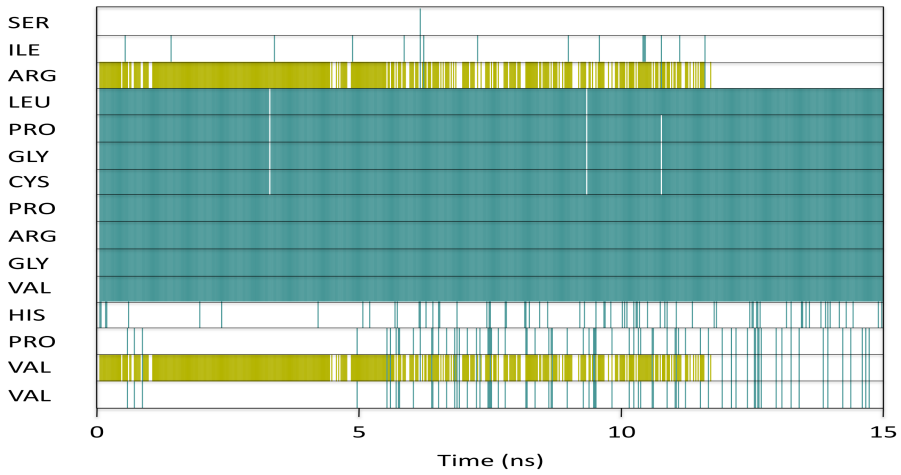


Figure 5.29: Secondary structure of each residue of the histidine mutant as a function of time. Green corresponds to turn, white is unordered and gold is β-strand.

were calculated over the last 5ns of the simulation (Fig. 5.30). The two central amino acids were found to have torsional angles (ϕ, ψ) equalling approximately -75°, -30° (Arg) and -85°, -10° (Gly), indicating a type I β-turn.

As in the hCGβ_{66-80} model, analysis of the bond probability distributions (Fig. 5.31) revealed the presence of two peaks - the location of one peak was found to be invariable irrespective of the chosen time frame (2.36Å), while the other varied to a minute degree (3.84-3.95Å). These peaks indicated that the hydrogen bond was indeed present during the course of the simulation due to the presence of the peak below the 2.5Å threshold. The presence of the hydrogen bond and type I β-turn in this system, would appear to support the notion that the ability of 8G5 to discriminate between the β sub-units of hCG and LH was due to a well defined structural difference between the two peptides.

In order to determine if the structure of the LHβ_{86-100} peptide was shared by other mutants (i.e. a type II β-turn), molecular dynamics simulations were performed on sequences which presented a lack in ability to successfully bind to the mAb 8G5. Analysis of the model representative of the sequence SIRLPGCPRGV**L**PVV revealed the peptide to present a type II β-turn over the Pro-Arg-Gly-Val sequence, as evident from the Ramachandran angles of these residues over the last 5ns of the simulation (Fig. 5.32). Furthermore, as with the model involving the histidine substitution, STRIDE predicted an increased stability in the turn motif spanning residues Leu-

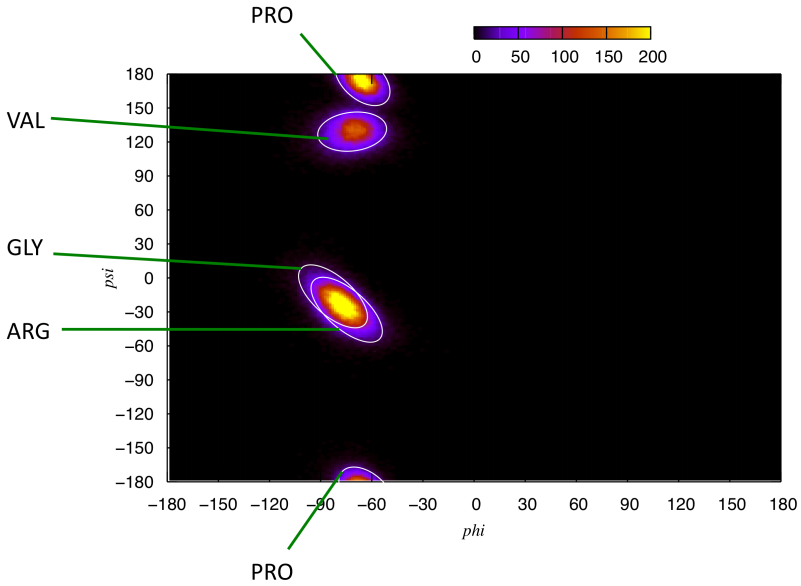


Figure 5.30: Ramachandran angles of sequence PRGV in the histidine mutant over the last 5ns. The torsional angles of the two central amino acids, Arg and Gly, indicate a type I β-turn.

Pro-Gly-Cys and in the β-sheet character between residues Arg and Val (Fig. 5.33); however as these features were not found to be stable in either hCGβ_{66-80} or LHβ_{86-100}, it was unlikely that they were important in rationalising the observed difference in binding affinity between these two antigens.

Another model investigated was representative of the sequence SIRLPGCPRGV**G**PVV. The structure of which was found to differ substantially from the other peptide sequences, with the model displaying 3_{10}-helical structure over the sequence RGVGPV (Fig. 5.34); this was confirmed by analysis of the torsional angles of Pro-Arg-Gly-Val, which revealed considerable intensity within the helical region of dihedral space (Fig. 5.35). As both peptide sequences (i.e. SIRLPGCPRGV**L**PVV and SIRLPGCPRGV**G**PVV) were found to be lack the ability to successfully bind to 8G5, the models would suggest that any deviation from the structure shared by hCGβ_{66-80} and the histidine mutant, would result in significant loss in binding; both SIRLPGCPRGV**L**PVV and SIRLPGCPRGV**G**PVV should lack the ability to form a hydrogen bond between the side chain of Leu or Gly, and the backbone amine group of Val.

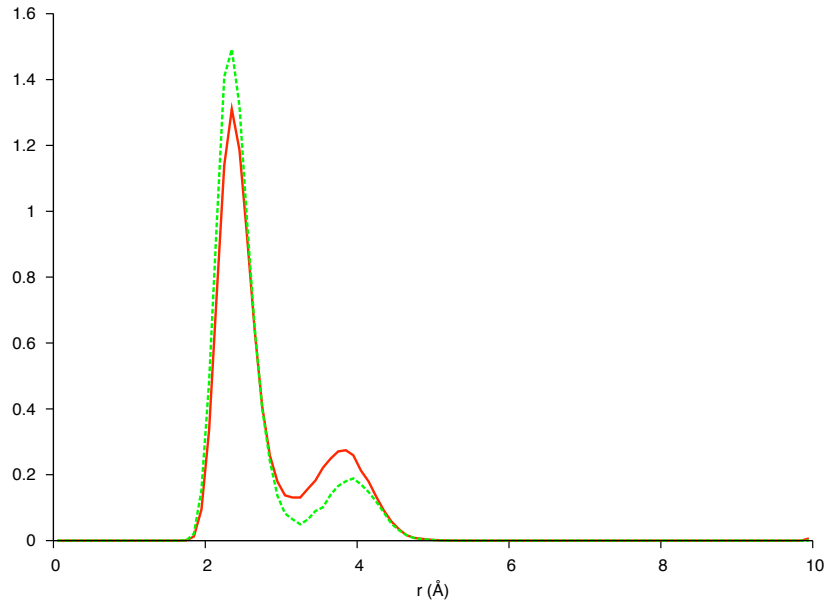


Figure 5.31: Bond probability analysis of the safety catch in the histidine mutant. Solid (red) line: over full 15ns, Dashed (green) line: over last 5ns.

An interesting question therefore was whether the presence of this hydrogen bond between the side chain of the reside located at position 77 and the backbone of Val_{79} (based on amino-acid numbering in hCGβ_{66-80}) stabilised a particular turn type over Pro-Arg-Gly-Val? Or, wether the residue at position 77 influenced the particular turn type, allowing for a spatial geometry accepting of such a hydrogen bond?

5.3.5 Protonation of Aspartate

In *Section 5.3.3.4*, it was hypothesised that the clasp between the side chain of Asp_{97} and the backbone amine group of Val_{99} would only be feasible if the Asp residue was protonated (Fig. 5.26c). Therefore it was of interest to investigate if such an interaction was possible and if the clasp had any bearing on the geometry of the turn experienced over the sequence Pro-Arg-Gly-Val. To that end, a simulation was set up in which one of the oxygen atoms (O_A^-) was protonated. This was akin to simulating the peptide under acidic conditions as the pKa (Eq. 5.1) of the aspartic acid side chain is 3.90 (Tab. A.1); K_a is the acid dissociation constant (Eq. 5.2).

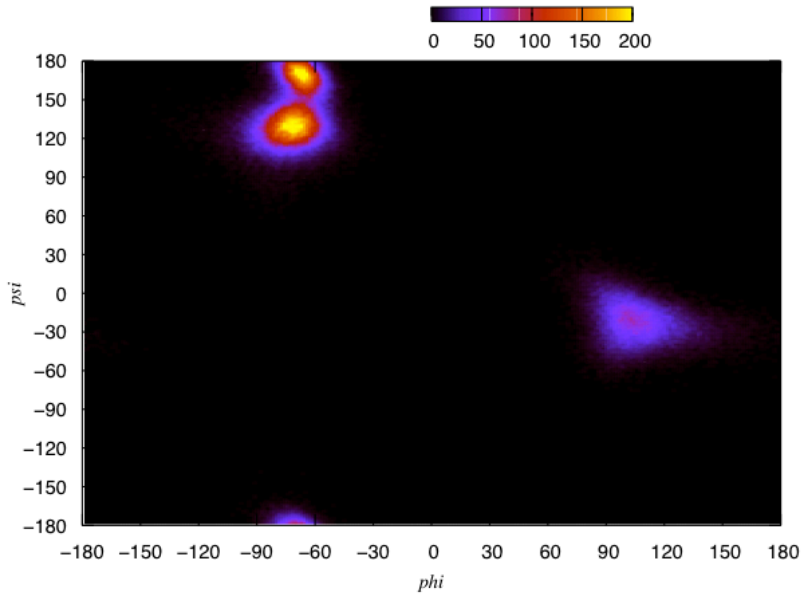


Figure 5.32: Ramachandran angles of sequence PRGV in the leucine mutant over the last 5ns. The torsional angles of the two central amino acids, Arg and Gly, indicate a type II β-turn.

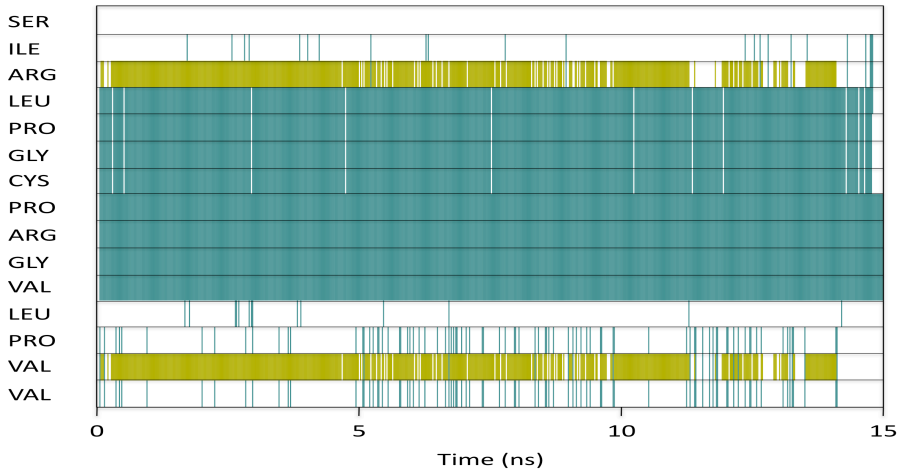


Figure 5.33: Secondary structure of each residue of the leucine mutant as a function of time. Green corresponds to turn, white is unordered and gold is β-strand.

From the Henderson-Hasselbalch equation (Eq. 5.3), a 50:50 ratio of deprotonated and protonated side-chains occurs at a pH value of 3.90, while a pH value of 1.90 equates to a ratio of 100:1 in favour of the protonated state (Eq. 5.4). Under such acidic conditions, one would expect the

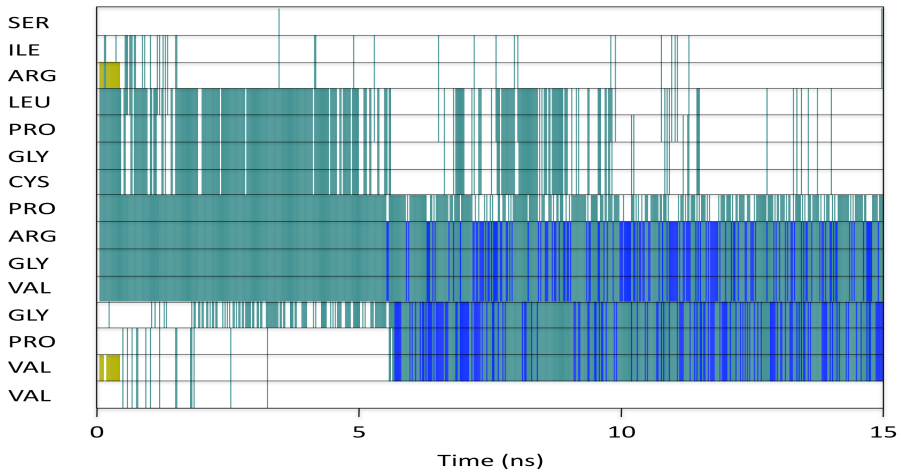


Figure 5.34: Secondary structure of each residue of the glycine mutant as a function of time. Green corresponds to turn, white is unordered and gold is β-strand.

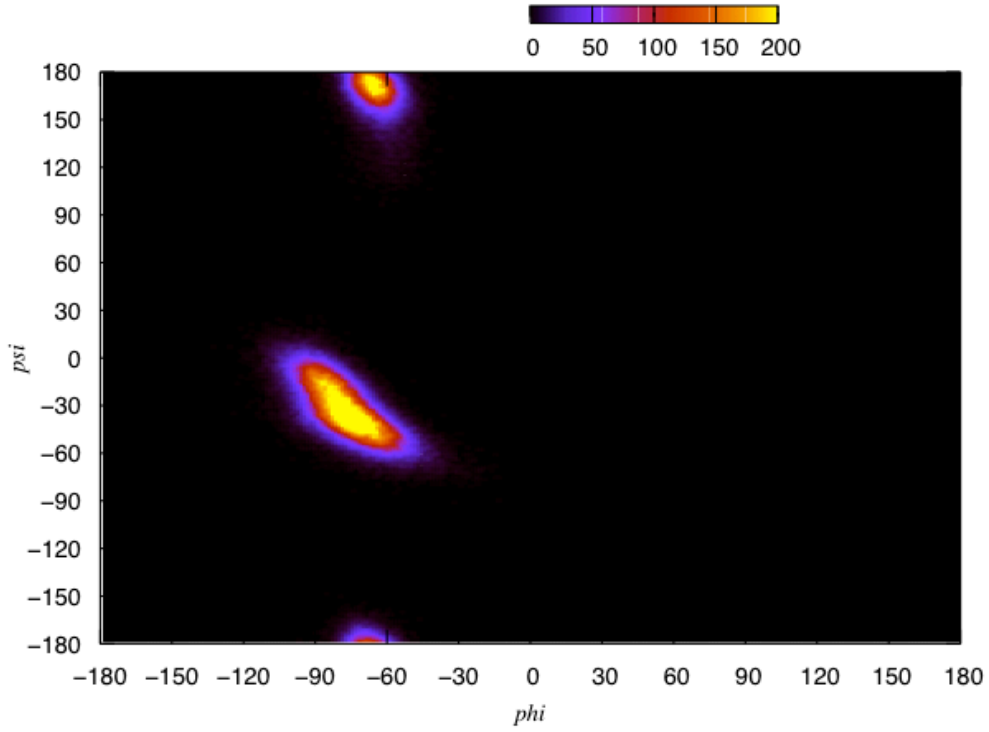


Figure 5.35: Ramachandran angles of sequence PRGV in glycine mutant over the last 5ns. The torsional angles, located within the helical region of dihedral space, support the 3_{10}-helical structure.

carboxylic acid group of the Val termini to also be partially protonated; therefore it is important to state that the conditions tested in the simulation reflect only a theoretical model.

$$pK_a = -log_{10}K_a \tag{5.1}$$

$$K_a = ([A^-][H^+]/[HA]) \tag{5.2}$$

$$pH = pK_a + log([A^-]/[HA]) \tag{5.3}$$

$$([A^-]/[HA]) = 10^{(pH-pKa)} \tag{5.4}$$

Analysis of the bond probability distribution for the carbonyl group of the protonated aspartic acid and backbone amine group of valine (Fig. 5.36) revealed that, over the last 5ns, only a single peak positioned at 5.958Å was present. Furthermore, over the full 15ns (and first 10ns), only minute traces were detectable below the assumed 2.5Å threshold of a hydrogen bond. This would suggest that the initial hypothesis (i.e. the protonated aspartic acid would be accepting of a hydrogen bond) was incorrect.

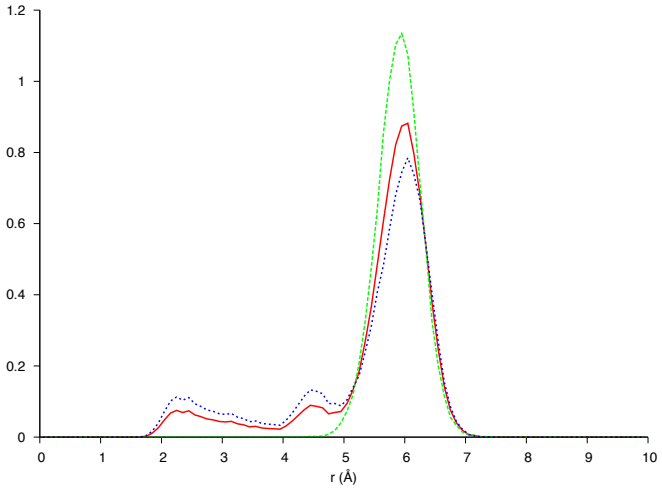


Figure 5.36: Bond probability analysis of the safety catch in the protonated aspartic acid mutant. Solid (red) line: over full 15ns, Dashed (green) line: over last 5ns, Dotted (blue) line: over initial 10ns.

As the simulation of this protonated form also predicted a stable turn over the sequence Pro-Arg-Gly-Val, for completeness, the Ramachandran angles were evaluated over the last 5ns of the simulation (Fig. 5.37). Interestingly the two central amino acids were found to have torsional angles equalling approximately -65°, -45° (Arg) and -80°, -20° (Gly). This is in stark contrast to those obtained for the deprotonated form of LHβ_{86-100} (Fig. 5.24), indicating that

the protonated form adopts a type I (or type III) β-turn as opposed to the type II β-turn found in the native LHβ_{86-100}.

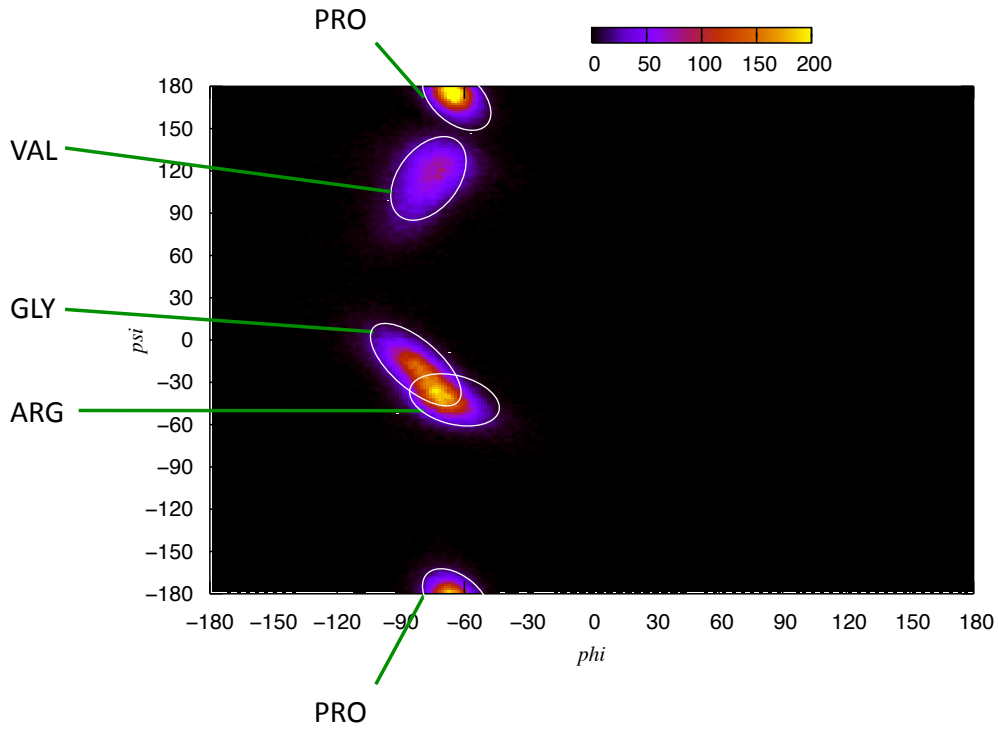


Figure 5.37: Ramachandran angles of sequence PRGV in protonated aspartic acid mutant over the last 5ns. The torsional angles of the two central amino acids, Arg and Gly, indicate a type I β-turn.

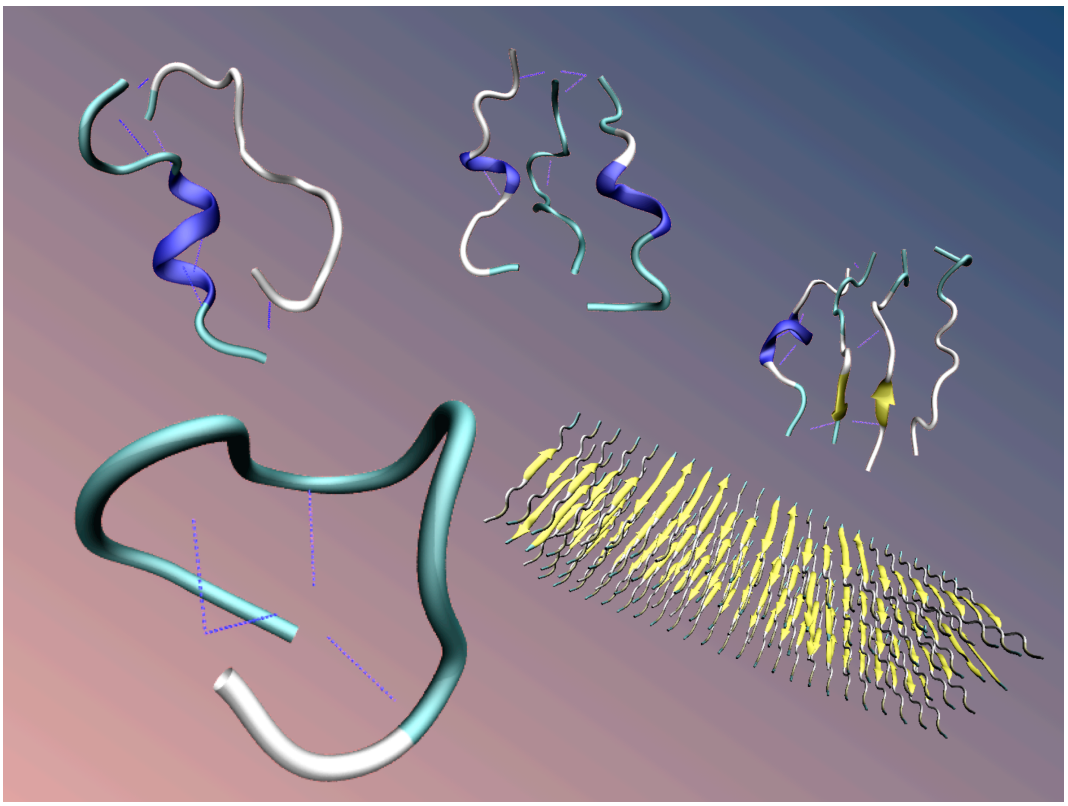
Such a result was surprising, as it was initially believed that the protonation would only affect the ability to form a hydrogen bond (which it did not) and would not have any impact on the structure immediately preceding the residue (which it did). In addition, the protonated model revealed that the two structural features (i.e. hydrogen bond and turn type) were not directly linked; the presence of a type I β-turn did not have to be accompanied by the hydrogen bond clasp. This therefore indicated that the presence of a particular turn type did not automatically allow for the spatial geometry accepting of such a clasp, and also suggested that the hydrogen bond clasp was not necessary for the stabilisation of a particular turn type. Whether the presence of both structural features was necessary for successful antigen-antibody binding remained unanswered.

5.4 Conclusions

The monoclonal antibody 8G5 is found to be able to discriminate between peptides representative of the β3 loops of hCGβ and LHβ, which are distinguishable by only a single amino acid mutation. Structural analysis of these peptides utilising molecular dynamics simulations reveals that while both sequences are characterised by a β-turn spanning the residues Pro-Arg-Gly-Val, the type of β-turn varies. In comparison to the results obtained experimentally, the secondary structures are consistent with an ensemble of open structures and turn motifs. The simulations of these peptides also reveal the presence of a proposed hydrogen bond clasp, present only in hCGβ_{66-80}. As this feature is present in the crystallographic structure of the hCGβ sub-unit, it would suggest that such a feature is important in the stabilisation of a particular turn type and thus the molecular recognition mechanism with the mAb 8G5.

6

Aggregation: Gas-phase transthyretin oligomers



6.1 Introduction

Self-assembly is regarded as a powerful approach for fabricating supramolecular architectures. In order for it to occur, it requires that the individual components be complementary to one another in shape. Furthermore, it is driven by the presence of numerous (individually) weak non-covalent interactions (e.g. hydrophobic interactions, electrostatic interactions and hydrogen bonds) [14]. Of the intermolecular forces, hydrophobic interactions are perhaps the most common driving force for the self-assembly of molecules in water, while electrostatic forces act on the largest length scale. Nature itself utilises self-assembly in order to generate larger, well-defined structures from building blocks such as DNA and peptides, which have themselves been the basis in building synthetic nano-structures [187, 188]. A number of applications have been suggested for materials prepared by the self-assembly of peptides, for example as scaffolding for tissue repair [189, 190], hydrogels and drug delivery vehicles [191]. As such, recent work in the field has concentrated on using the self-assembly of small peptides to generate new materials. In addition, there has been significant research into the self-assembly of peptides in an effort to understand the formation of amyloid fibrils; small peptide fragments of larger proteins prone to such amyloid formation have been extensively used as models for the proteins themselves.

Amyloid fibrils are constructed from proteins that have converted from their soluble native structure to insoluble fibrils which form either systemically or within specific organs such as the brain or liver. Many human and animal disorders such as Alzheimer's, Parkinson's, Creutzfeldt-Jakob, bovine spongiform encephalopathy, kuru, and type II diabetes are associated with the formation of these amyloid fibres [23, 24, 192–194]. Amyloid fibrils are not the native conformation of the protein, but would appear to be a thermodynamically favourable alternative structure that is formed spontaneously. Furthermore, numerous proteins and peptides that are not connected to any known disease have been shown to form amyloid fibrils *in vitro* [192, 194–197], indicating that the ability to form such fibrils may be an inherent property of peptides [195].

Interestingly, proteins lacking structural or functional similarities have been found to self-assemble into amyloid fibrils which display similar chemical and physical properties [198]. In amyloid-like structures, the individual polypeptide chains form β-strands that lie perpendicular to the long axis of the fibre. As indicated by their distinctive X-ray pattern, protofilaments

consist of a cross-β-sheet structure, with the distance separating adjacent sheets falling between 10-15 Å, while neighbouring polypeptide chains within the sheets are separated by 4.7Å; it is these structures that subsequently self-assemble to give amyloid fibrils [199]. Other commonly shared characteristics [192, 194] are that the fibrils are typically long, unbranched and $\sim$100Å in diameter (as revealed by electron microscopy), and bind to the dye Congo red, resulting in a green birefringence under polarised light.

6.2 System chosen

Human transthyretin (TTR) is a 55-kDa protein involved in the transport of thyroxine and retinol in plasma, and the native protein is a homotetramer of 127-residue subunits with extensive β-sheet structure [200]. The wild-type protein, as with many of its fragments and naturally occurring mutants, self-assembles into amyloid fibres *in vivo* and *in vitro* which are associated with diseases including senile systemic amyloidosis (SSA) and familial amyloid polyneuropathy (FAP) [201, 202]. In addition, two 11-residue peptide fragments derived from the native sequence, designated TTR_{10-20} and $\text{TTR}_{105-115}$ (Fig. 6.1), have been shown to form amyloid fibrils *in vitro* [203]. These peptides represent β-strands A and G of the full-length TTR and are both located at the surface of the thyroxine-binding channel formed by the TTR homotetramer. Due to their ability to form fibrils which exhibit the characteristic 'cross-β' x-ray diffraction pattern [204], these peptides are ideal candidates as model systems in order to investigate the structure of amyloid fibrils and the mechanism of their formation. This is exemplified by previous investigation into the structure of $\text{TTR}_{105-115}$ in amyloid fibrils utilising nuclear magnetic resonance (NMR) [205, 206], FTIR [207] and molecular dynamics simulations focusing on the mechanism of fibril formation [208–210]. Time-course mass spectrometry (MS) and ion-mobility mass spectrometry (IM-MS) have also been used to probe the early aggregation states of the $\text{TTR}_{105-115}$ peptide [209], and have provided insight into the 3-dimensional geometry of the various oligomers formed during the initial stages of self-assembly by means of calculated collision cross-sections. The experimental results were compared to calculated values from simulations of multimers of charge $(n-1)+$ (where n denoted the number of TTR peptides) which were achieved by incorporating an anionic C-terminal carboxyl to one of the $\text{TTR}_{105-115}$

peptides; an exception was a simulation of the monomeric species which consisted of a cationic N-terminus and neutral C-terminus giving rise to a net charge of 1+.

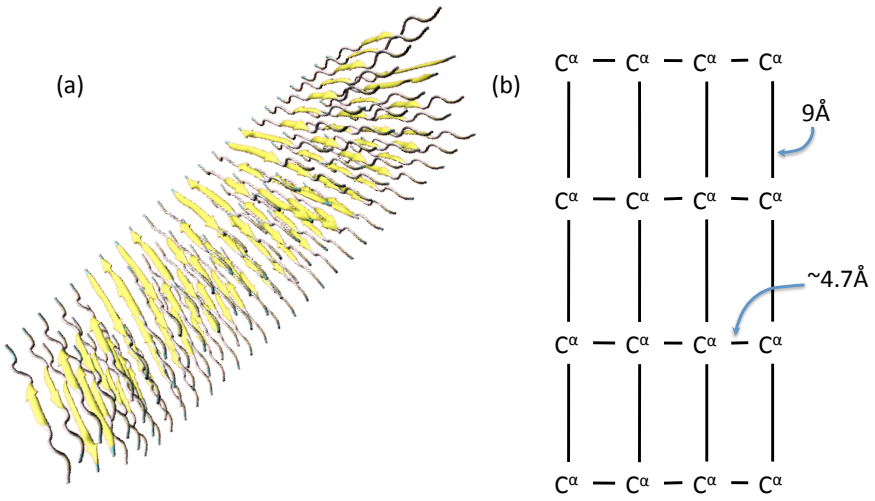


Figure 6.1: (a) Visualisation of the theoretical model of the amyloid structure of $TTR_{105-115}$, revealing four stacked antiparallel β-sheets (antiparallel with respect to the peptides within each sheet and between neighbouring sheets) [5], and (b) illustration of the separation between the C^{α} in Leu_{110}, revealing the distance between sheets to be 9Å, while the distance between peptides within the same sheet is 4.7Å

It was of interest to probe whether gas-phase simulations of neutral $TTR_{105-115}$ monomers and oligomers resulted in structures and subsequent collision cross-sections comparable to those of the charged species. Such an investigation was hoped to reveal whether the charge state influenced the packing of the oligomers and the energy landscape of the TTR molecule as it aggregated.

6.3 Computational Methods

Two unit cells were set up, which each comprised one TTR molecule (Fig. 6.2a) with zwitterionic termini in a cubic box of length 80.0Å, with the peptide being described by either the CHARMM22 [31] or AMBER-ff03 [36, 37] force fields. From the theoretical model of the amyloid structure (Fig. 6.1), four antiparallel β-sheets are stacked upon one another; each peptide is antiparallel with respect to one another within each sheet, and between neighbouring sheets. While moving onto dimeric and higher order oligomers, it was therefore important to consider various permutations that such systems could adopt; the dimer could be comprised of two mo-

lecules from the same sheet, or two from different sheets. Two unit cells were subsequently set up, which comprised either two TTR molecules from the same sheet (Fig. 6.2b) or two from neighbouring sheets (Fig. 6.2c) in a cubic box of length 80.0Å. These cells were then replicated and the peptides were handled by the CHARMM22 [31] or AMBER-ff03 [36, 37] force fields.

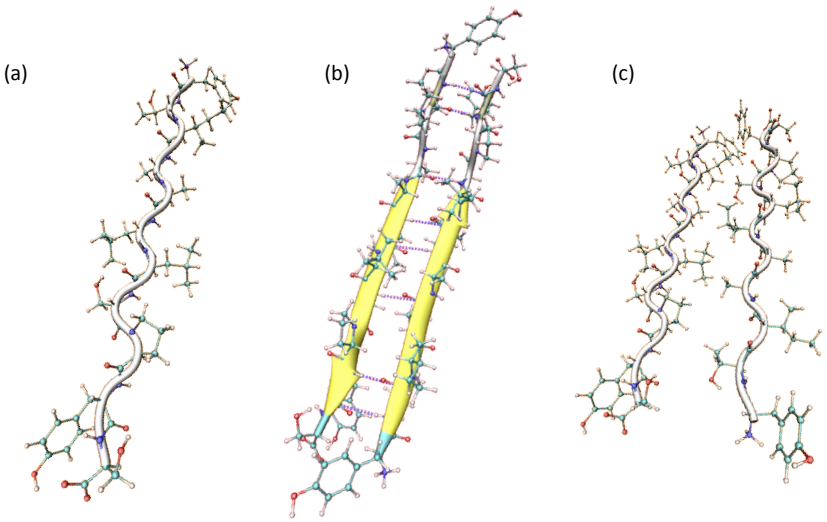


Figure 6.2: Snapshot of the initial structure of the $TTR_{105-115}$ (a) monomer, (b) intra-sheet dimer and (c) inter-sheet dimer.

A similar approach was applied to a trimeric system of TTR molecules, with unit cells being created which consisted of either three molecules from the same sheet (Fig. 6.3a) or from neighbouring sheets (Fig. 6.3b), while each system was represented by either the CHARMM22 or AMBER-ff03 force field (all within a cubic box of length 80.0Å).

In the case of the tetramer system, in addition to two unit cells comprised of four TTR molecules from either the same sheet (Fig. 6.4a) or from neighbouring sheets (Fig. 6.4b), a further unit cell comprised of two intra-sheet dimer systems forming a rectangular cuboid shape (Fig. 6.4c). As with the other systems, these unit cells were replicated and were handled by either the CHARM22 or AMBER-ff03 force fields.

Each of the 16 simulations (8 systems, 2 force fields) were equilibrated at 300K in the canonical ensemble for 650ps with a time-step of 0.5fs to anneal out unphysical contacts. The system was then run for a further 150ps in the canonical ensemble employing a time-step of 1fs, before a production run of 5ns using a time-step of 1fs from which all data was collected. Periodic boundary conditions were employed and long-range interactions were evaluated via Ewald

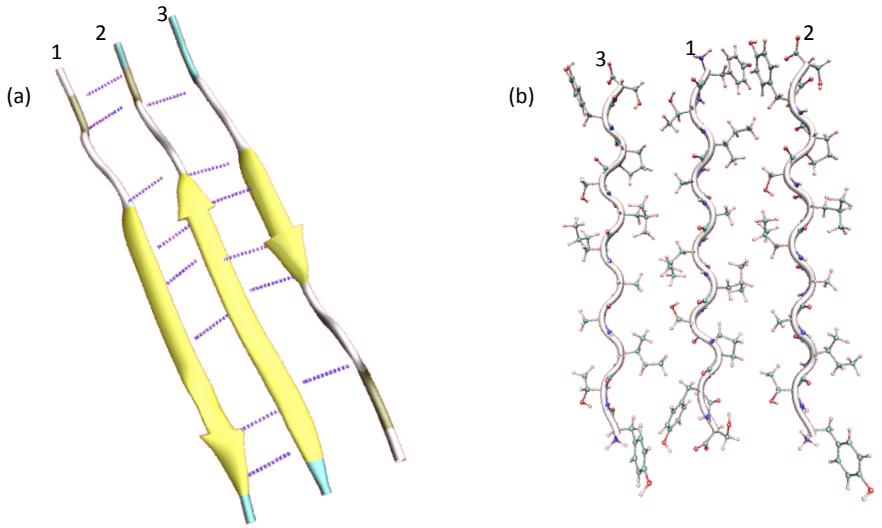


Figure 6.3: Snapshot of the initial structure of the $TTR_{105-115}$ (a) intra-sheet trimer and (b) inter-sheet trimer, coupled with peptide designations (i.e. Peptide 1, Peptide 2 etc.).

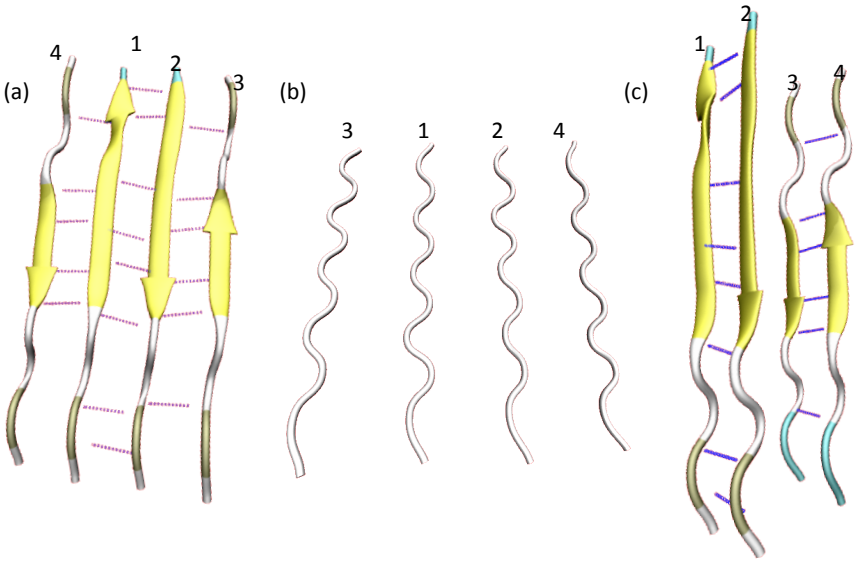


Figure 6.4: Snapshot of the initial structure of the $TTR_{105-115}$ (a) intra-sheet, (b) inter-sheet and (c) 'cuboid' tetramer, coupled with peptide designations (i.e. Peptide 1, Peptide 2 etc.).

summation. The PINY simulation package [99] was employed to perform the simulations, while collision cross-sections were estimated with *mobcal* using the trajectory method [211]. This method treats species as a collection of atoms (each represented by a Lennard-Jones and an ion-induced dipole potential) in order to obtain a scattering angle representative of the angle between the incoming and departing buffer gas atom trajectory. The averaged collision integral

is subsequently determined by averaging over all possible collision geometries.

6.4 Results

Due to previous experiments being conducted utilising mass spectroscopy, it was important to investigate the structure of possible oligomers detected utilising this technique. To that end, a series of MD simulations of gas phase peptide systems were conducted; the gas phase was chosen to mimic the vaporised analyte. It should be noted that while mass spectroscopy makes use of ionised particles, these simulations were conducted on zwitterionic, whole peptides. This was in aid of comparison with simulations conducted by another group in parallel, which focused on simulating the ionised gas phase molecules; such a comparison was hoped to probe whether the structure and size of the ionised peptide was retained in the neutral free peptide.

6.4.1 Monomeric Transthyretin

In order to determine if the structure of a monomer of TTR remained extended in the gas phase, two simulations were performed utilising the CHARMM22 and AMBER-ff03 force field; the use of two separate force fields would expose any pronounced sensitivity to standard empirical potentials. Visualisation of the simulations after 5ns revealed that the initially extended rod-like appearance of the monomer had collapsed into a compact structure under both parameterisations (Fig. 6.5). In support of this, analysis of the average end-to-end distance over the last 1ns (defined as the distance between the N atom of the Tyr_{105} N-terminus and the C atom of the Ser_{115} C-terminus) revealed sharp peaks located at 3.14Å and 2.95Å for the CHARMM22 and AMBER-ff03 parameterisations respectively. In comparison to MD simulations conducted on monomeric $TTR_{105-115}$ in solution, such a distance was found to range from 3Å to 33Å with an average of 14.2Å [208]. Clearly the gas phase simulations conducted herein resulted in a significantly more collapsed geometry; this was not surprising as the system was comprised of only a single molecule in the gas-phase, and the simulation would lack any stabilising intermolecular interactions.

To ascertain whether their were subtle differences between the two systems, the secondary structure of each residue as a function of time was analysed utilising the STRIDE secondary

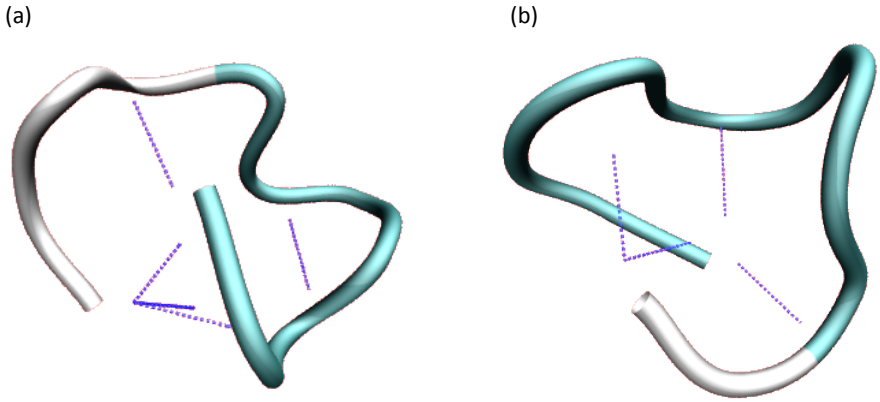


Figure 6.5: Visualisation of the structure of the $TTR_{105-115}$ monomer after 5ns, under the (a) CHARMM22 and (b) AMBER-ff03 parameterisation. In relation to the initial structure, the peptide backbones were found to present RMSD values of 9.244Å and 9.528Å respectively.

structure assignment algorithms [107] (Fig. 6.6, 6.7). From this, it was clear that both simulations predicted stable turn-like motifs spanning the sequence Tyr_{105}-Leu_{111} over the course of the 5ns, with the AMBER-ff03 parameterisation extending this span to include the Ser_{112} and Pro_{113} residues; the CHARMM22 force field also displayed turn structure over these residues, however this was only a transient feature. In addition, the AMBER-ff03 force field revealed additional secondary structure, with transient appearances of β-sheet character between the Ala_{108} and Ser_{112} residues, and 3_{10}-helical structure spanning the sequence Ala_{109}-Leu_{111}.

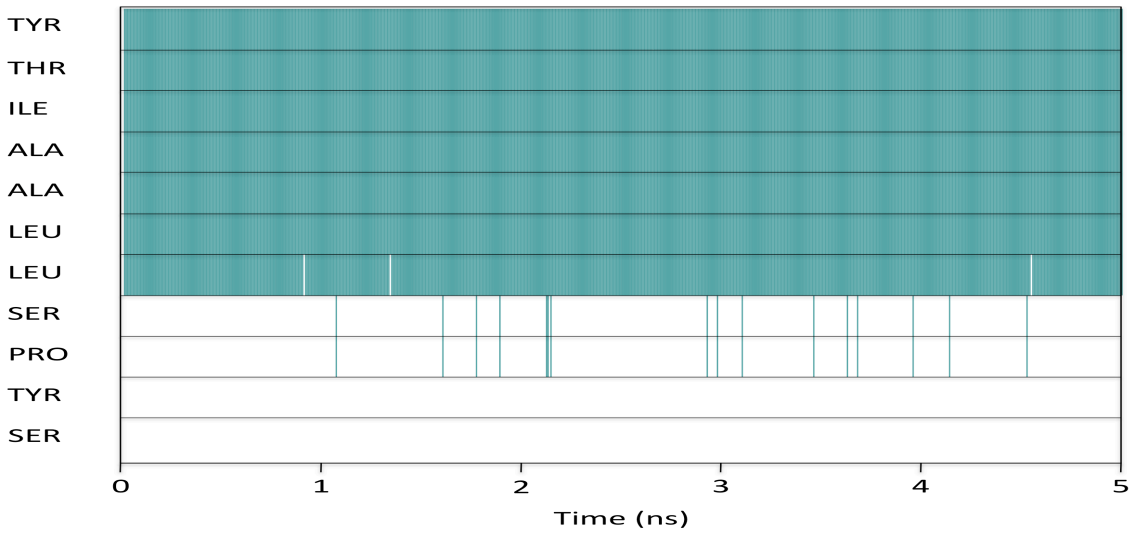


Figure 6.6: Secondary structure of each residue of $TTR_{105-115}$ under the CHARMM22 parameterisation as a function of time. Green corresponds to turn and white is unordered. Simulation predicts stable turn motifs over Tyr_{105}-Leu_{111}.

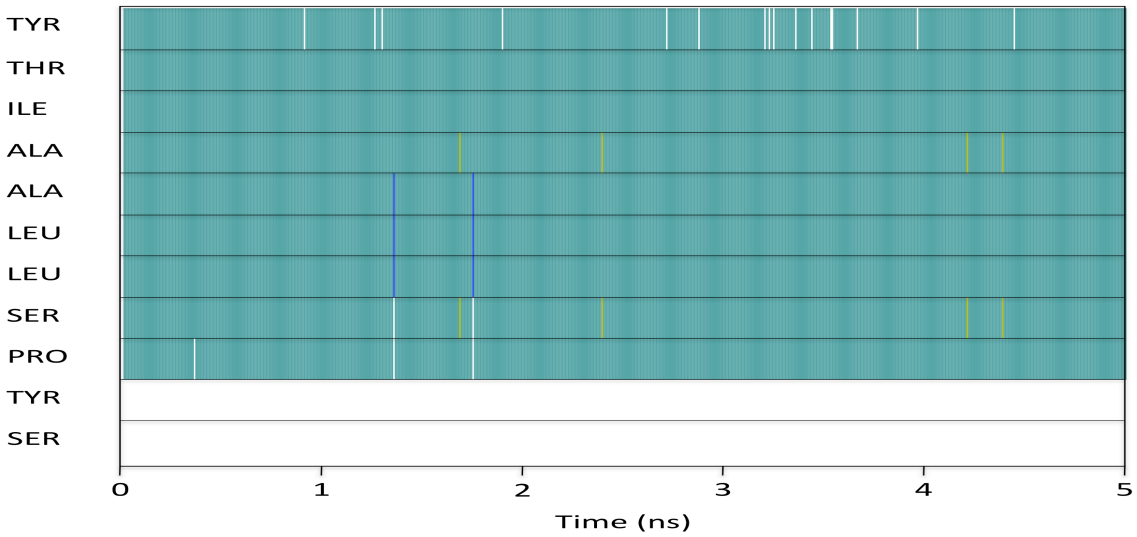


Figure 6.7: Secondary structure of each residue of $TTR_{105-115}$ under the AMBER-ff03 parameterisation as a function of time. Green corresponds to turn, white is unordered, blue is 3_{10}-helix and gold is β-strand. Simulation predicts stable turn motifs over Tyr_{105}-Pro_{113}.

With both simulations predicting the monomeric gas phase system to adopt similar secondary structure and similar end-to-end distances, it was important to determine if the 3-dimensional structures occupied similar volumes or if one simulation predicted a more compact structure. To that end, the simulations were used to calculate the cross-sectional area of the peptides (Fig. 6.8). From the data over 5ns, both the CHARMM22 and AMBER-ff03 parameterisations, predicted the monomeric TTR peptide to adopt a structure with a similar cross-sectional area; over the last 1ns, CHARMM22 resulted in an average of 273Å^2, while AMBER-ff03 resulted in an average of 270Å^2; the experimentally obtained value of the collision cross-section of $[M$+H$]^+$ was found to be 256Å^2 [209]. The only notable difference between the two simulations was that the AMBER-ff03 force field predicted the monomer to adopt a cross-sectional area of approximately 260Å^2 for almost 1ns, indicating a more compact geometry over this time period.

It is therefore reassuring that from the collected data, there did not appear to be any significant force-field bias in the handling of the intramolecular interactions of the gas phase TTR, as evident from both systems predicting similar 3-dimensional architecture. In comparison to previous simulations conducted on the monomeric TTR [209], all three MD simulations were found to present near identical collision cross-sections. Furthermore, the AMBER-ff03 simulation was found to transiently lie within the bounds of the experimental uncertainty (256$\pm$6Å^2)

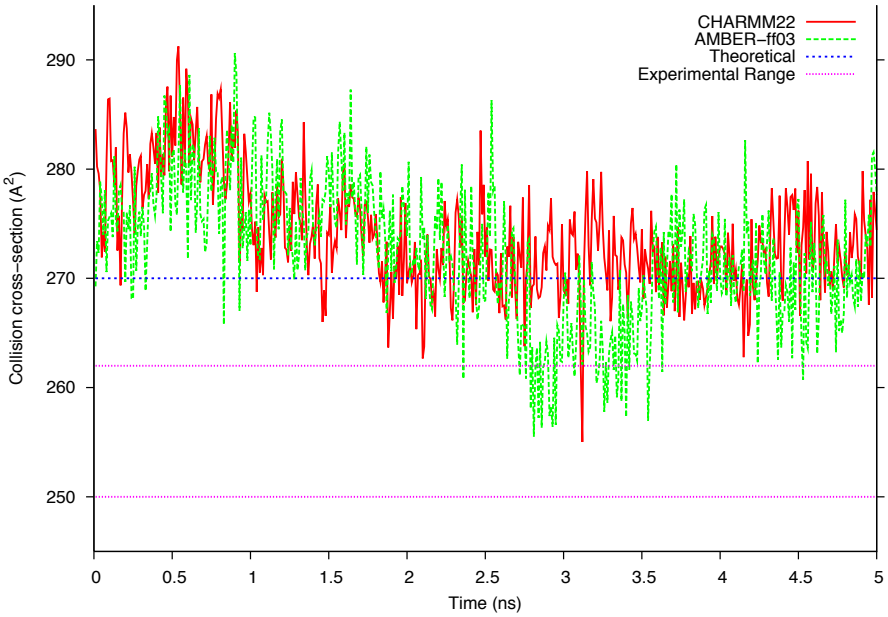


Figure 6.8: Collision cross-sections of the monomer simulations as a function of time, compared to the previously reported values of $[M$+H$]^+$ obtained experimentally and via simulation.

suggesting that the structure measured utilising IM-MS was a compact structure similar in structure to that of the simulations (Fig. 6.5); the average collision cross-sections over the last 1ns were found to deviate from the experimental mean by 5.47% and 6.64% for the AMBER-ff03 and CHARMM22 parameterisations respectively.

6.4.2 Dimeric Transthyretin

Due to the predicted structure of TTR consisting of four stacked β-sheets, it was of interest to probe both a dimeric system from the same sheet ("intra-sheet dimer") and two peptides taken from neighbouring sheets ("inter-sheet dimer"). For the purpose of comparing secondary structure, it was decided to ignore the labels of the peptides in the dimeric system (i.e. Peptide 1 and Peptide 2). This was based on the assumption that as the two identical antiparallel peptides would be exposed to the same environment, their order of input in defining the simulation would have no bearing on their secondary structure evolution.

6.4.2.1 Intra-sheet dimer

Visualisation of the intra-sheet dimer system after 5ns revealed that under the CHARMM22 parameterisation, one of the two peptides had adopted a 3_{10}-helix , while the other had formed sequential turns and remained fairly extended resulting in the system forming a 'β-bulge'-like structure (Fig. 6.9a). STRIDE revealed that the 3_{10}-helix predominantly spanned the sequence Ala_{108}-Leu_{111} (Fig. 6.10) as evident from this structure being relatively stable over the course of the 5ns; however this sequence did appear to transiently adopt an α-helix, and was also sporadically assigned turn-like structure (which extended to include Ile_{107}-Ser_{112}). Analysis utilising STRIDE also revealed the predominantly turn-like peptide to adopt occasional 3_{10} architecture over the Ile_{107}-Ala_{109} and Pro_{113}-Ser_{115} sequences.

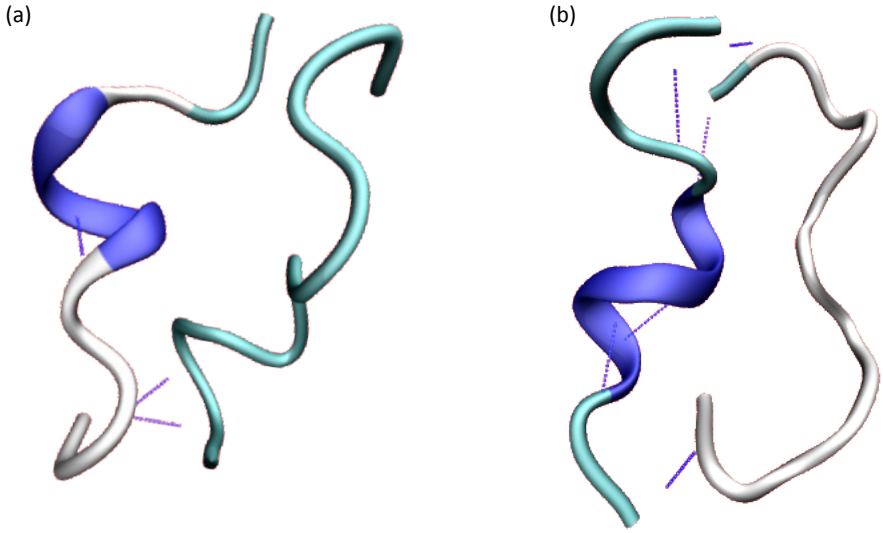


Figure 6.9: Visualisation of the structure of the $TTR_{105-115}$ intra-sheet dimer after 5ns, under the (a) CHARMM22 and (b) AMBER-ff03 parameterisation. In relation to the initial structure, the peptide backbones were found to present RMSD values of 6.243Å and 6.010Å respectively.

Visualisation of this dimeric system under the AMBER-ff03 parameterisation similarly revealed one peptide to adopt 3_{10}-helical structure, while the other had formed an extended structure (Fig. 6.9b). However, analysis of the secondary structure of each residue as a function of time exposed notable differences (Fig. 6.11). One peptide from each system was indeed defined by a 3_{10}-helical structure over the last 2ns of the simulations, with the span slightly longer under the AMBER-ff03 force field: Ile_{107}-Leu_{111} (*cf.* Ala_{108}-Leu_{111} under CHARMM22). The non-helical peptide was found to adopt a predominantly random coil structure over the last 3ns,

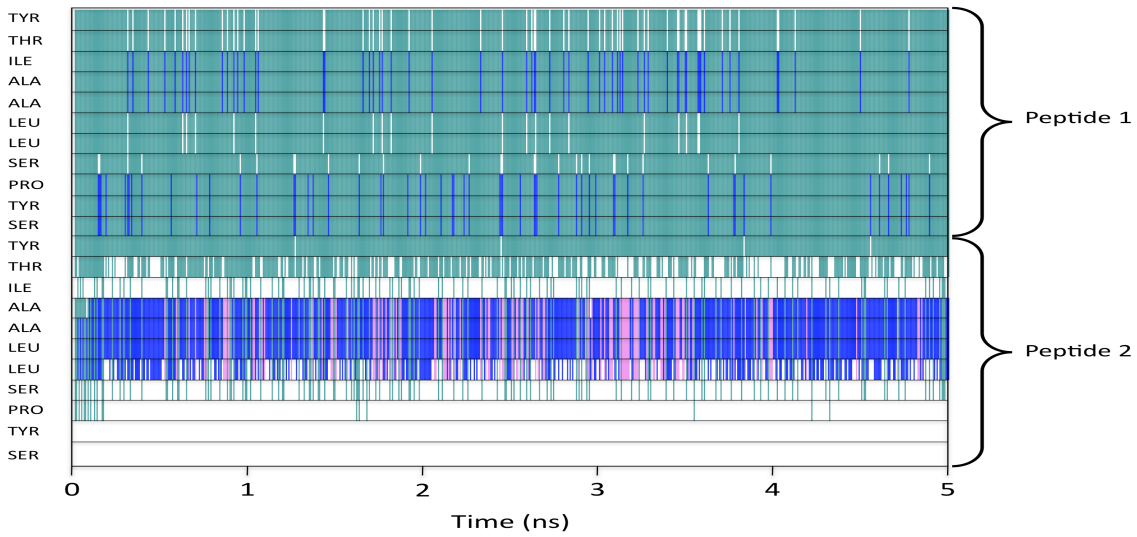


Figure 6.10: Secondary structure of each residue of the intra-sheet dimeric $TTR_{105-115}$ under the CHARMM parameterisation as a function of time. Green corresponds to turn, white is unordered, blue is 3_{10}-helix and pink is α-helix.

wheres under the CHARMM22 parameterisation the non-helical peptide was assigned turn-like structure; under both parameterisations, the non-helical peptide did remain fairly extended. However it was in the initial 2ns where the two simulations differed the most. During this time, the AMBER-ff03 parameterisation predicted the initial β-sheet character to persist, particularly between the Leu_{110}-Leu_{111} residues of one peptide and the Ala_{109}-Leu_{110} residues of the other.

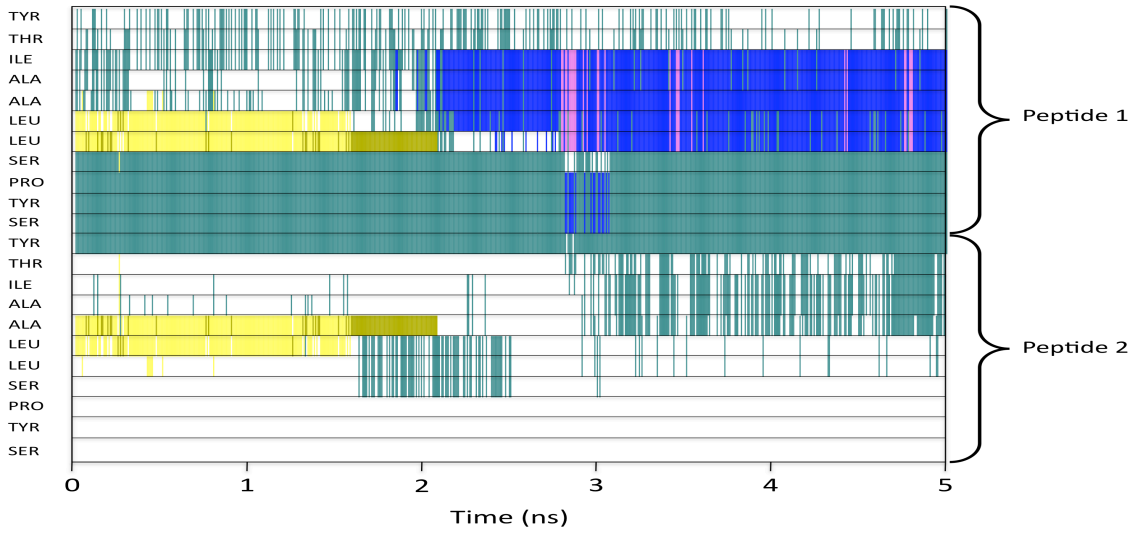


Figure 6.11: Secondary structure of each residue of the intra-sheet dimeric $TTR_{105-115}$ under the AMBER-ff03 parameterisation as a function of time. Green corresponds to turn, white is unordered, blue is 3_{10}-helix, pink is α-helix, yellow is β-sheet and gold is β-strand.

Having revealed a sensitivity to standard empirical potentials in intermolecular interactions, it was interesting to probe how such a sensitivity affected the cross sectional area exhibited by the peptide systems; from the analysis utilising STRIDE it was safe to assume that the peptide system had adopted a stable state by 3-4ns under both parameterisations, and it was anticipated by which point the collision cross-section should have reached an equilibrium. While the CHARMM22 force field resulted in a steady collisional cross-section of 460-470Å^2 (Fig. 6.12), the production run of the AMBER-ff03 system was found to initiate from a significantly larger geometry (540Å^2). Comparing the plotted collisional cross-section and the secondary structure of each residue over time, a correlation between the loss in β-sheet character (in the AMBER-ff03 system) and the fall in collision cross-section could be detected. As both systems resulted in a similar structure and collisional cross-section (466Å^2 and 462Å^2 for the CHARMM22 and AMBER-ff03 force fields respectively over the last 2ns), it would appear that the β-bulge structure (observed under both parameterisations over the last 2ns) was more compact than the initial β-sheet dimer.

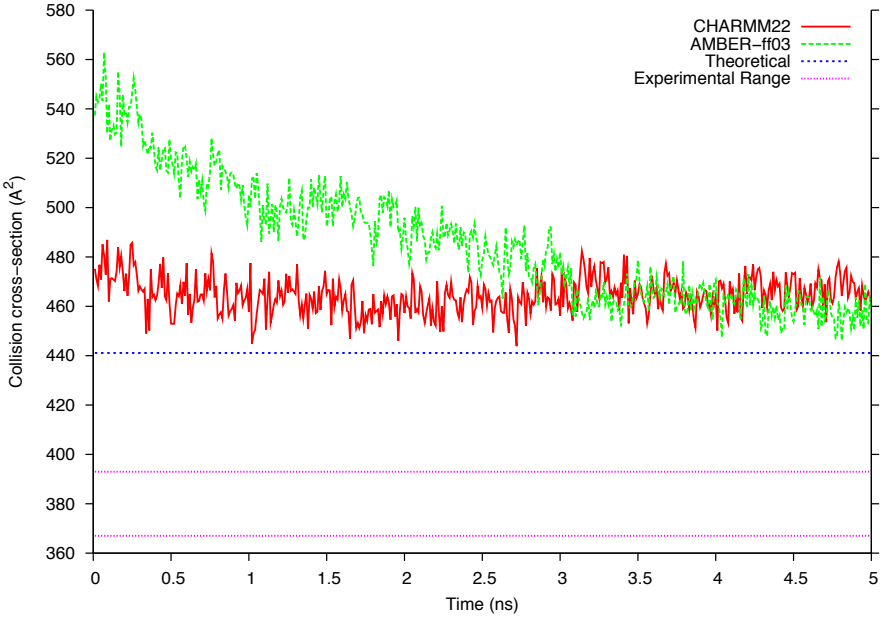


Figure 6.12: Collision cross-sections of the intra-sheet dimer simulations as a function of time, compared to the previously reported values of $[2M$+2H$]^+$ obtained experimentally and the $[2M$+H$]^+$ simulation.

Previous simulations of the various oligomers predicted a collision cross-section of 441.1Å^2 for the $[2M$+H$]^+$ species [209], a difference of 5.2% in comparison to the simulations con-

tained herein (assuming an average 464Å^2); these two values corresponded to diameters equating 21Å and 21.54Å respectively. The recorded data therefore indicated that the simulations of the neutral dimer, resulted in structures comparable to those reported utilising charged species. Therefore, introduction of a single H^+ did not seem to affect the structure of the system dramatically. However in comparison to the collision cross-sections obtained experimentally, there was clearly a significant difference between simulation (those contained herein and reported previously) and experiment. The experimentally recorded cross-section area (380$\pm$13Å^2) suggested that the geometry of the species recorded utilising IM-MS was significantly more compact than what any of the three simulations predicted; the simulations were found to result in a collision cross-section area 22.1% greater than that obtained experimentally. It should be noted that in the case of the dimeric system, the previously reported collision cross-sections represented two different ionic species ([2M+H]$^+$ for simulation *cf.* [2M+2H]$^{2+}$ for experiment), which may indicate that the difference in charge state did indeed affect the structuring of the dimeric system; however, the values for the neutral and [2M+H]$^+$ simulations were similar. The increase in charge would be expected to expand the structure and thereby increase the collision cross-sectional area as a result of increased electrostatic repulsion. From the data collected herein, the neutral species were clearly found to result in far larger species than one would have therefore expected.

6.4.2.2 Inter-sheet dimer

Visualisation of the inter-sheet dimeric system after 5ns revealed the peptides to adopt similar arrangements as those of their intra-sheet counterparts (Fig. 6.13). Both force fields resulted in each TTR molecule adopting helical structure: under the AMBER-ff03 parameterisations both TTR molecules adopted 3_{10}-helical structure, while the CHARMM22 force predicted one to adopt a 3_{10}-helix and the other an α-helix.

Analysis of the secondary structure of each residue over the last 2ns revealed that under both parameterisations (Fig. 6.14, 6.15) the two peptides were predominantly defined by helical regions spanning the Ala_{108}-Leu_{111} and Ile_{107}-Leu_{111} sequences. While the Ile_{107}-Leu_{111} sequence of one peptide was principally an α-helix over the 5ns production run of the CHARMM22 system, the AMBER-ff03 force field predicted the structure to initiate from a 3_{10}-helix spanning

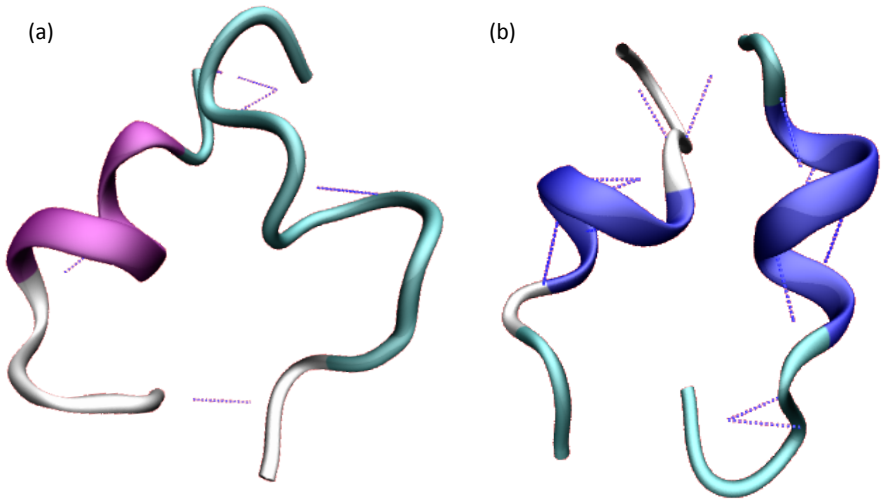


Figure 6.13: Visualisation of the structure of the $TTR_{105-115}$ inter-sheet dimer after 5ns, under the (a) CHARMM22 and (b) AMBER-ff03 parameterisation. In relation to the initial structure, the peptide backbones were found to present RMSD values of 6.831Å and 6.229Å respectively.

the Ile_{107}-Ala_{109} residues, adopting α-helical content as the simulation evolved. Conversely, the AMBER-ff03 parameterisation predicted the Ala_{108}-Leu_{111} sequence of the other peptide to adopt a stable 3_{10}-helix over the simulation, while the STRIDE secondary structure algorithm assigned this sequence in the CHARMM22 system to be initially turn-like over the initial 2ns before adopting a mixture of α- and 3_{10}-helix.

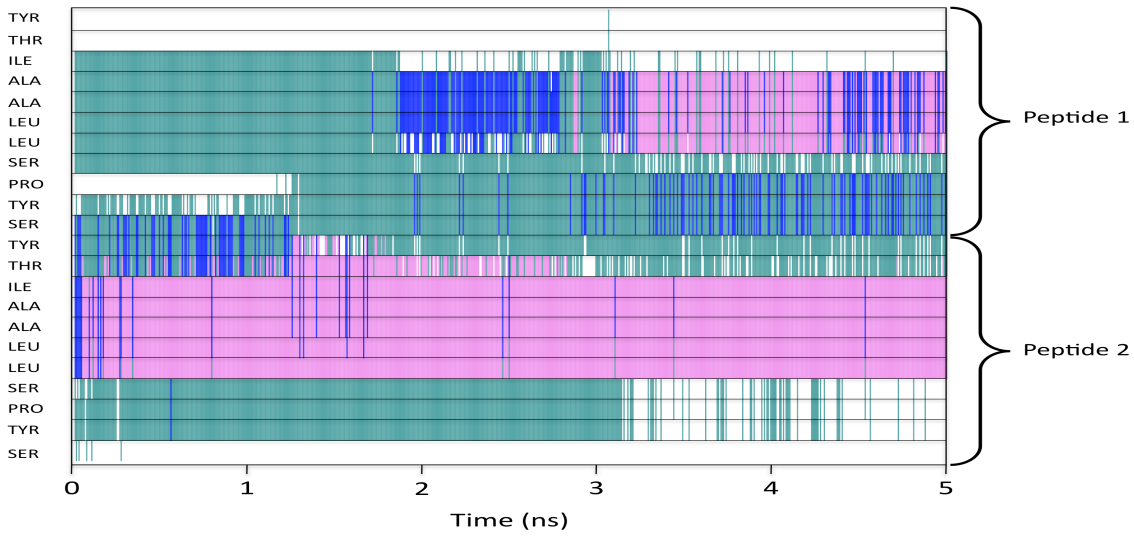


Figure 6.14: Secondary structure of each residue of the inter-sheet dimeric $TTR_{105-115}$ under the CHARMM parameterisation as a function of time. Green corresponds to turn, white is unordered, blue is 3_{10}-helix and pink is α-helix.

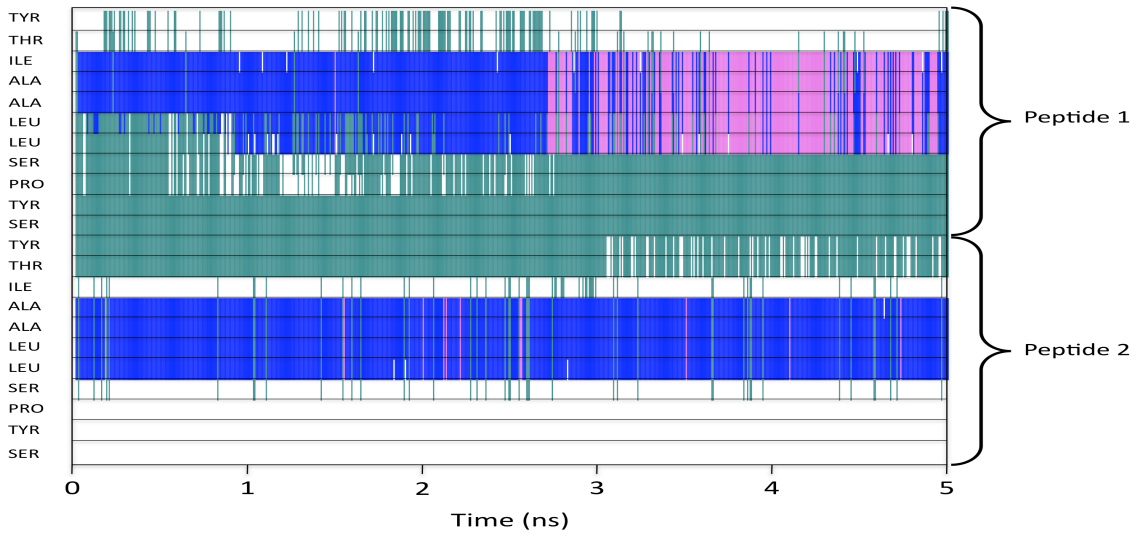


Figure 6.15: Secondary structure of each residue of the inter-sheet dimeric $TTR_{105-115}$ under the AMBER-ff03 parameterisation as a function of time. Green corresponds to turn, white is unordered, blue is 3_{10}-helix and pink is α-helix.

Both parameterisations resulted in similar collision cross sections over the course of the 5ns simulation with no observable force field bias (Fig. 6.16). Furthermore, the values of these were comparable to those achieved by the intra-sheet dimer system over the last 2ns, with the CHARMM22 and AMBER-ff03 force fields resulting in values of 460Å^2 and 458Å^2 respectively. This suggested that without the support of initial β-sheet interactions, the dimer system instantly collapsed into the more compact structure similar in size to that eventually observed in the intra-sheet dimer. These collision cross sections were found to result in deviations of approximately 4% and 21% from the previous simulation and experimentally measured values.

6.4.3 Trimeric Transthyretin

As with the dimeric system, two trimer systems of TTR were investigated comprising of either 3 peptides from the same β-sheet ("Intra-sheet trimer") or 3 from neighbouring sheets ("Inter-sheet trimer"). In the dimeric system, the peptide labels were ignored due to the symmetry of the system. While this may hold true for the two outermost peptides in the trimer (i.e. Peptide 1 $\leftrightarrow$ Peptide 3) as both peptides are in close proximity with only one other TTR molecule, the middle peptide must be treated independently. Therefore for simplification and anticipation of peptide drifts (changing the order of the strands), all amino acids will be labelled with an index

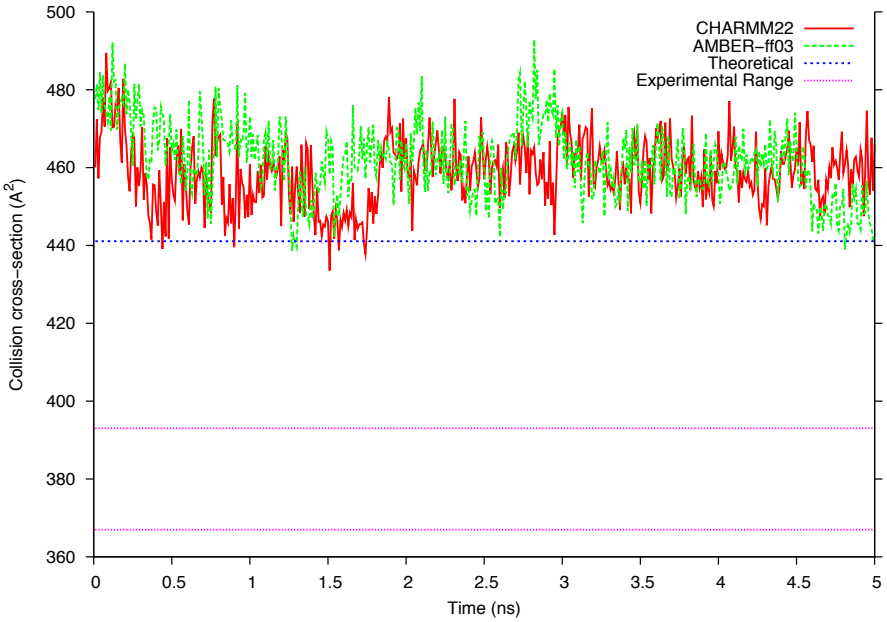


Figure 6.16: Collision cross-sections of the inter-sheet dimer simulations as a function of time, compared to the previously reported values of $[2M+2H]^+$ obtained experimentally and the $[2M+H]^+$ simulation.

denoting their chain (i.e. $\text{Pro}(1)_{113}$, $\text{Pro}(2)_{113}$ or $\text{Pro}(3)_{113}$) in order to compare interactions between strands. Additionally, a shorthand notation '$\angle$**a.b**' shall be used to denote the angle between the end-to-end vectors of peptide 'a' and peptide 'b' (i.e. $\angle$**1.2**, for peptides 1 and 2).

6.4.3.1 Intra-sheet trimer

After 5ns, the CHARMM22 force field predicted the trimeric system to remain fairly extended exhibiting a slight curve (Fig. 6.17a). Analysis of the end-to-end distances revealed an average length of 20.3̇Å, while angles of 177.7° ($\angle$**1.2**) and 174.1° ($\angle$**2.3**) were found between these end-to-end vectors, indicating that the strands remained aligned: assuming 145° as a minimum cut-off for an antiparallel alignment [208]. The model also demonstrated evidence of inter-molecular hydrogen bonding between the neighbouring strand's backbones. Evidence of β-sheet characteristics was also observed between the $\text{Leu}(1)_{110}$ & $\text{Leu}(2)_{110}$ and $\text{Pro}(1)_{113}$ & $\text{Ile}(2)_{107}$ pairings. Visualisation of the AMBER-ff03 system, however, revealed the outermost peptides had adopted partial 3_{10}-helical content while the central TTR molecule had formed a large turn-like hairpin motif (Fig. 6.17b). The end-to-end distances were found to be smaller than under

the CHARMM22 parameterisation, resulting in an average 17.93̇Å, however the peptides were found to remain fairly aligned: 178.2° (∠**1.2**) and 167.5° (∠**2.3**).

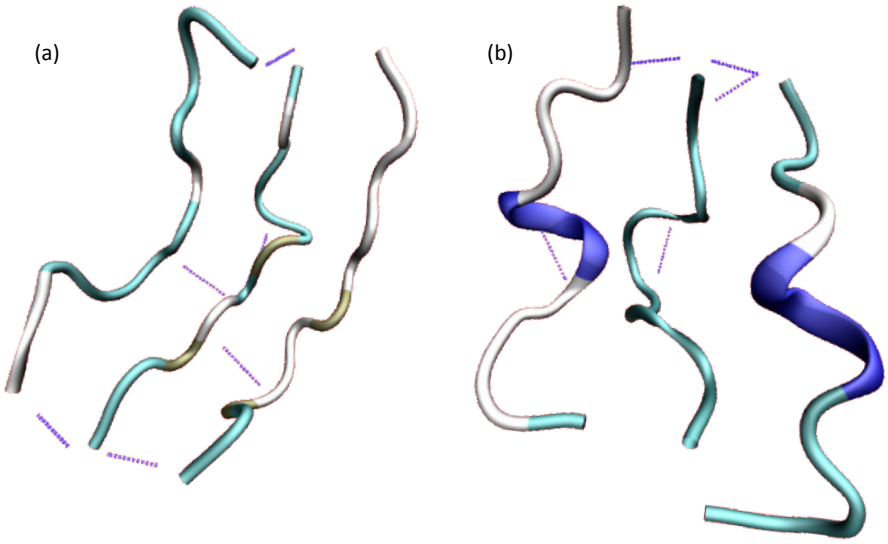


Figure 6.17: Visualisation of the structure of the $TTR_{105-115}$ intra-sheet trimer after 5ns, under the (a) CHARMM22 and (b) AMBER-ff03 parameterisation. In relation to the initial structure, the peptide backbones were found to present RMSD values of 5.399Å and 6.317Å respectively.

Analysis of the secondary structure of each residue in the CHARMM22 system as a function of time revealed β-sheet character in the $Leu(1)_{110}$, $Pro(1)_{113}$, $Ile(2)_{107}$ and $Leu(2)_{110}$ residues over the entire course of the simulation (Fig. 6.18). The $Ile(1)_{107}$, $Pro(2)_{113}$ and $Leu(3)_{110}$ residues were also assigned such structure over the initial 2ns. From the orientations of the peptides (i.e. the antiparallel arrangement), the stable β-sheet interactions must be from the pairings of $Pro(1)_{113}$ & $Ile(2)_{107}$, and $Leu(1)_{110}$ & $Leu(2)_{110}$; this was confirmed via visualisation of the system (Fig. 6.17a). Furthermore, over the initial 2ns, the $Ile(1)_{107}$ and $Pro(2)_{113}$ residues were also found to be aligned with one another in order to facilitate β-sheet interactions. The assignment of β-sheet interactions involving proline residues was a surprising result, as this residue was known to disrupt both β-sheets and α-helices (Tab. A.4); however prolines may be found in the edge strands of β-sheets. As the proline residues lacked a hydrogen in the amide group (Tab. A.1), such β-sheet character must have been as a result of the proline carbonyl group acting as a hydrogen bond acceptor. Finally, the $Leu(3)_{110}$ residue appeared to interact with the $Leu(2)_{110}$ residue, indicating that over the initial 2ns all three Leu_{110} residues in the trimeric system were perfectly aligned to form intermolecular interactions stabilising a β-sheet structure.

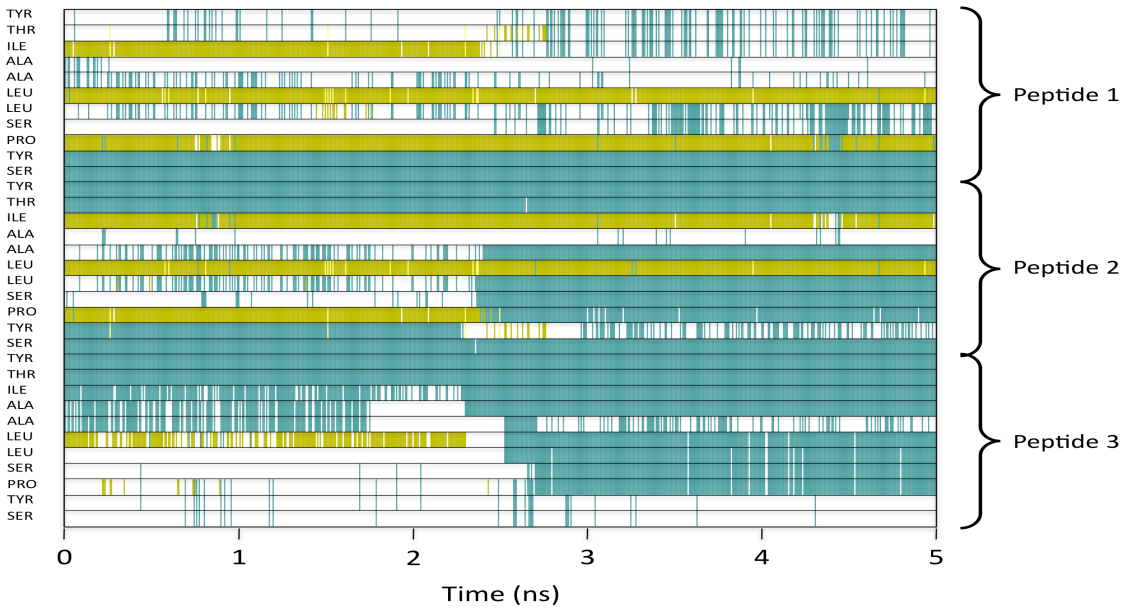


Figure 6.18: Secondary structure of each residue of the intra-sheet trimeric $TTR_{105-115}$ under the CHARMM parameterisation as a function of time. Green corresponds to turn, white is unordered and gold is β-strand.

The secondary structure of each residue over the course of the AMBER-ff03 simulation was also analysed (Fig. 6.19). Interestingly, the $Pro(1)_{113}$ & $Ile(2)_{107}$ interaction observed under the CHARMM22 parameterisation was also found to be stable over the initial 4ns of the AMBER-ff03 simulation. β-sheet interactions between the $Ala(1)_{108}$ and $Ser(2)_{112}$ were also observed up to 3.5ns into the simulation, which occasionally extended to include the sequences spanning $Ile(1)_{107}$-$Ala(1)_{109}$ and $Leu(2)_{111}$-$Pro(2)_{113}$, while there was no observed β-sheet character in the third peptide at all; over the initial 4ns this third peptide was found to adopt a mixture of random coil and turn-like structure (over Ala_{109}-Ser_{112}). Finally, over the course of the last 1ns, the two outermost peptides (designated Peptide 1 & 3) were assigned 3_{10}-helical structure over the Ala_{109}-Leu_{111} sequences. The initial stabilisation of β-sheet character and ultimate helix formation was comparable in both the dimeric and trimeric intra-sheet systems under the AMBER-ff03 parameterisations. In the investigation into force field bias on the predicted secondary structure of $gp41_{659-671}$ in *Chapter 3*, it was the CHARMM22 force field which was found to result in the greater helical composition. Therefore it was expected that if the peptide did indeed have helical propensity, of the two force fields employed, it would have been the CHARMM22 system to fold into such a structure.

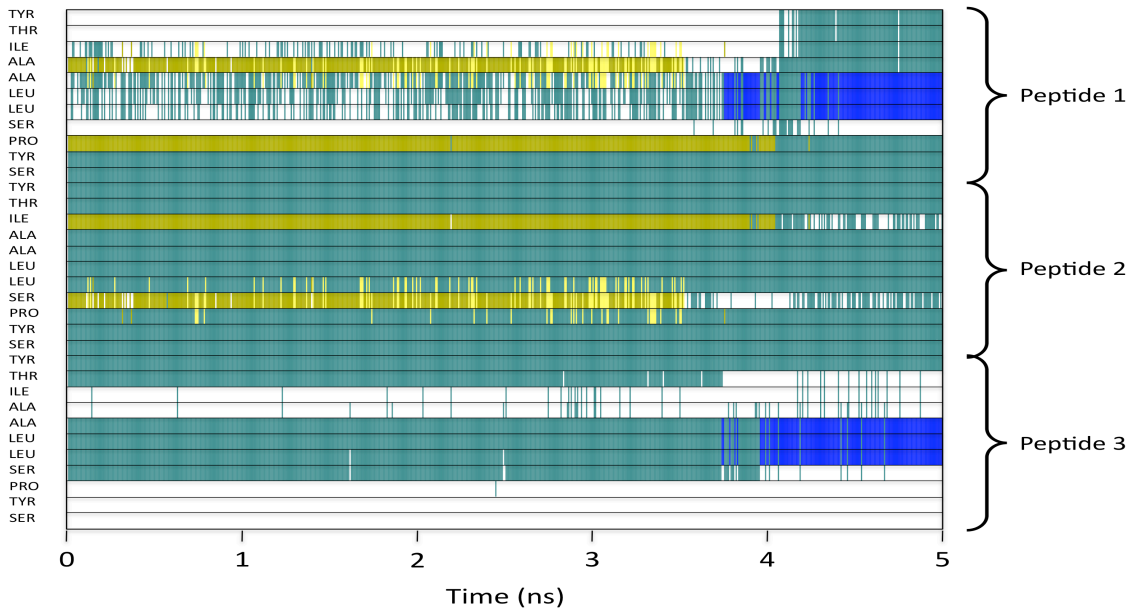


Figure 6.19: Secondary structure of each residue of the intra-sheet trimeric $TTR_{105-115}$ under the AMBER-ff03 parameterisation as a function of time. Green corresponds to turn, white is unordered, gold is β-strand and blue is 3_{10}-helix.

In the dimeric system, retained β-sheet character was found to correlate with a higher collision cross section (Fig. 6.12), therefore it was anticipated that there would be a notable difference in the trimeric system over the last 1ns due to the differing secondary structure assigned over this period from the different parameterisations. From the calculated collision cross section of the two simulations, the CHARMM22 system was found to decrease from an average area of approximately 675Å^2 to 600Å^2 over a period spanning the 2-3ns interval (Fig. 6.20). This correlated with the loss of β-sheet character involving the $Ile(1)_{107}$, $Pro(2)_{113}$ and $Leu(3)_{110}$ residues. Having started from a geometry of 680-700Å^2, the AMBER-ff03 system was found to collapse to approximately 640Å^2 until 3.5ns into the simulation. At which point, the loss of β-sheet character between the $Ala(1)_{108}$ and $Ser(2)_{112}$ residues coupled with the formation of 3_{10}-helical motifs, reduced the geometry of the trimeric system to 615Å^2 over the last 1ns - effectively matching the size predicted by the CHARMM22 parameterisation (607Å^2 over the same time period). Therefore, while the structures of the last 1ns of each simulation differed, the overall size of each system appeared to be consistent.

Previous MD simulations of a trimeric system $[3M+2H]^{2+}$ resulted in a theoretical collision cross-section of 567Å^2 [209]. The simulations of the neutral intra-sheet trimer were therefore

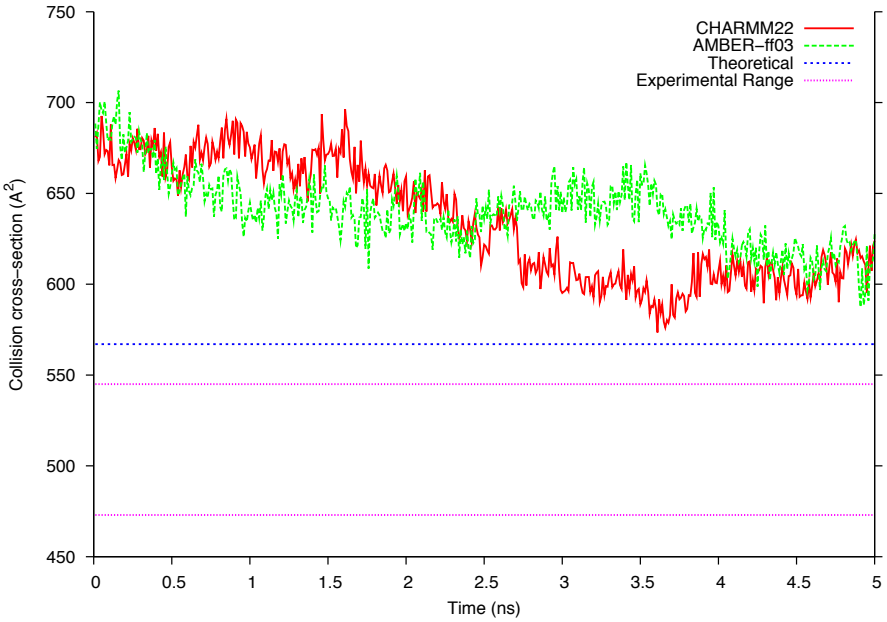


Figure 6.20: Collision cross-sections of the intra-sheet trimer simulations as a function of time, compared to the previously reported values of $[3M+2H]^{2+}$ obtained experimentally and via simulation.

found to result in areas approximately 7.05% and 8.47% (CHARMM22 and AMBER-ff03 respectively) larger than their charged counterparts. In comparison to the experimentally obtained value for the $[3M+2H]^{2+}$ trimer (509±36Å^2), the neutral species were found to deviate by 19.3% and 20.8%, indicating that the species measured experimentally adopted a more compact structure than any of the simulations predicted. The correlation of larger area and β-sheet character under the AMBER-ff03 and CHARMM22 parameterisations would suggest that such secondary structural features were not present in the true structure of $[3M+2H]^{2+}$.

6.4.3.2 Inter-sheet trimer

The CHARMM22 force field was found to result in the trimeric system to be defined by an α-helix flanked by extended/turn-like peptides (Fig. 6.21a), while under the AMBER-ff03 parameterisation, visualisation of the system revealed the 3 peptides to have collapsed into stacked 3_{10}-helices (Fig. 6.21b). Analysis of the end-to-end distances revealed that the trimer of peptides remained more extended in the AMBER-ff03 system resulting in an average length of 18.6Å compared to the average 13.36Å for CHARMM22; under the CHARMM22 force field, peptide 3

was found to remain more extended (16.6Å) than the other two peptides (12.2Å & 11.3Å), however all peptides were found to be significantly shorter than their original starting length of ∼35.7Å and their intra-sheet counterparts. Both force fields were found to result in the peptides remaining fairly aligned with one another; CHARMM22 resulted in end-to-end vector angles of 174.0° (∠**1.2**) and 162.0° (∠**2.3**), while AMBER-ff03 resulted in angles of 167.8° (∠**1.2**) and 154.3° (∠**2.3**), all above the 145° cut-off [208].

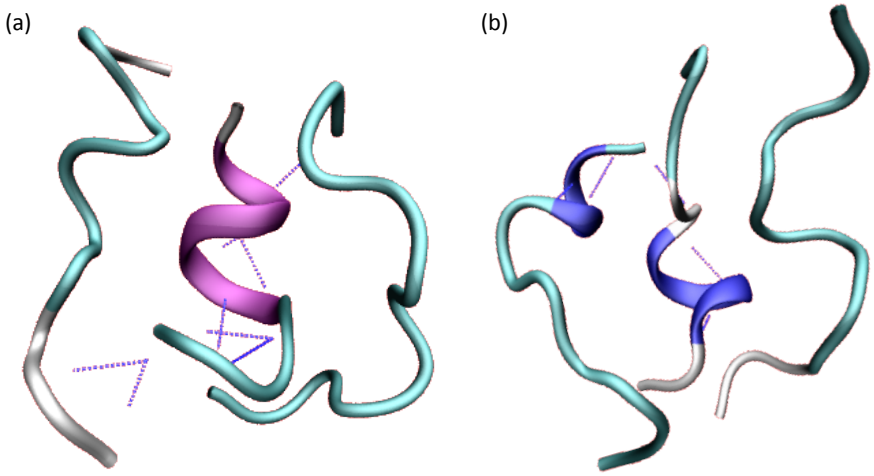


Figure 6.21: Visualisation of the structure of the $TTR_{105-115}$ inter-sheet trimer after 5ns, under the (a) CHARMM22 and (b) AMBER-ff03 parameterisation. In relation to the initial structure, the peptide backbones were found to present RMSD values of 7.425Å and 7.044Å respectively.

The α-helical motif was found to be a very stable motif over the course of the 5ns simulation under the CHARMM22 parameterisation, and was found to consistently span the $Thr(1)_{106}$-$Leu(1)_{111}$ sequence (Fig. 6.22). A 3_{10}-helical motif was also assigned transiently to the sequence $Ala(3)_{109}$-$Leu(3)_{111}$, less so as the simulation progressed. Analysis of the secondary structure of each residue in the AMBER-ff03 system revealed that the 3_{10}-helices spanned slightly different sequences in each of the peptides: $Ile(1)_{107}$-$Leu(1)_{110}$, $Thr(2)_{106}$-$Ala(2)_{109}$ and $Ala(3)_{108}$-$Leu(3)_{111}$.

In comparison to the inter-sheet dimer, both systems predicted stable α-helical structure over the Ile_{107}-Leu_{111} sequence under the CHARM22 force field, while the AMBER-ff03 handling of the systems predicted helical content over the Ile_{107}-Leu_{110} and Ala_{108}-Leu_{111} sequences. Such a result may seem unsurprising as the AALL sequence should present a high affinity for helical formation (Tab. A.4), and without the presence of initial β-sheet structure, TTR molecules may

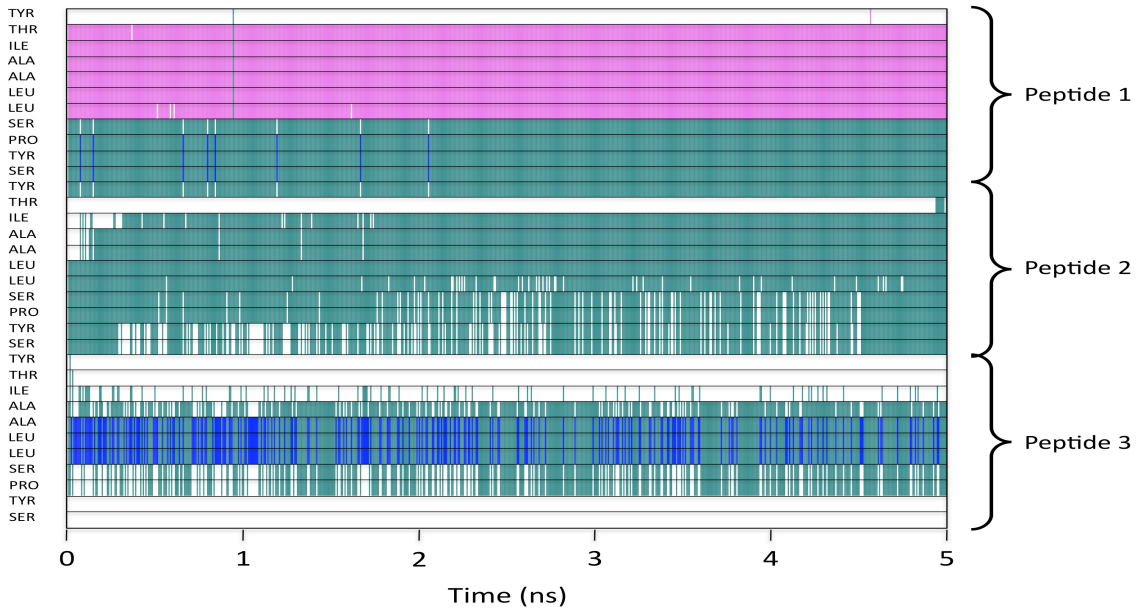


Figure 6.22: Secondary structure of each residue of the inter-sheet trimeric $TTR_{105-115}$ under the CHARMM parameterisation as a function of time. Green corresponds to turn, white is unordered, pink is α-helix and blue is 3_{10}-helix.

fold into such motifs. However as evident in both monomeric simulations (Fig. 6.5), the gas phase $TTR_{105-115}$ did not fold into a helix, instead it was found to collapse into a compact geometry of turn-like secondary structure.

In analysing the collision cross-sections over the course of the 5ns, both simulations revealed transitions from larger to smaller volumes (Fig. 6.24). The transition of an average area of approximately 620Å^2 to 600Å^2 in the AMBER-ff03 simulation after 2ns appeared to coincide with the assignment of turn motif over the $Ser(2)_{112}$-$Ser(2)_{115}$ sequence (Fig. 6.23). However the transition from 600Å^2 to 580Å^2 in the CHARMM22 simulation after 3ns did not appear to coincide with any change in secondary structure. Over the last 1.5ns, the two simulations resulted in collisional cross-sections of 585Å^2 (CHARMM22) and 606Å^2 (AMBER-ff03), coinciding with deviations from the previous simulated value by 3.2% and 6.9%, and deviations from the experimentally obtained value by 14.9% and 19.1%.

6.4.4 Tetrameric Transthyretin

In the previous dimeric and trimeric systems, two models were investigated which comprised molecules from either the same sheet or neighbouring sheet ("Intra-sheet" and "Inter-sheet" re-

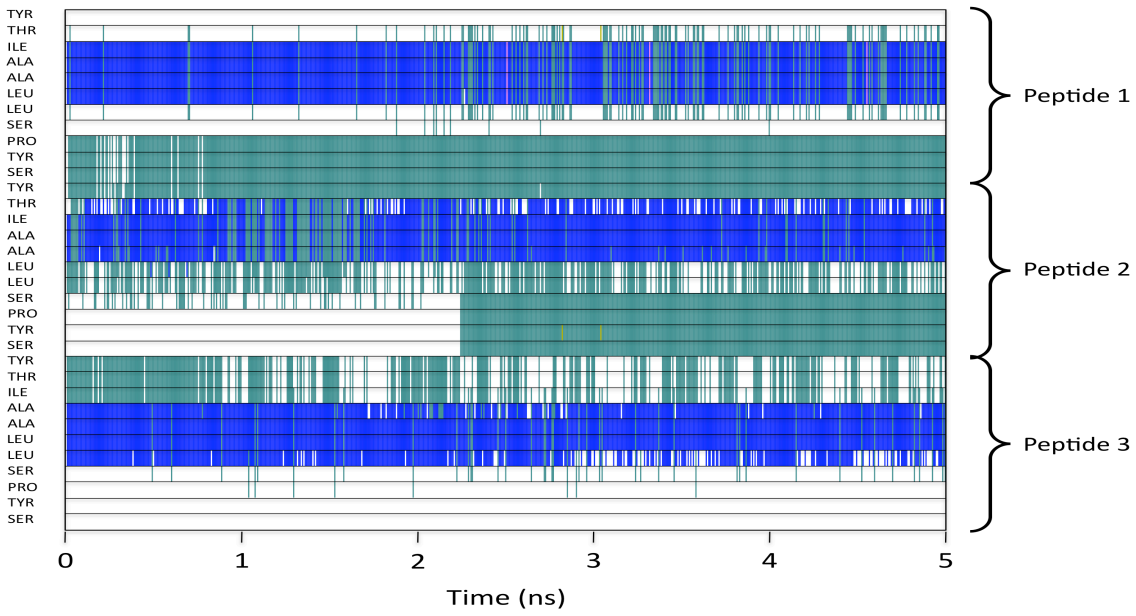


Figure 6.23: Secondary structure of each residue of the inter-sheet trimeric $TTR_{105-115}$ under the AMBER parameterisation as a function of time. Green corresponds to turn, white is unordered and blue is 3_{10}-helix.

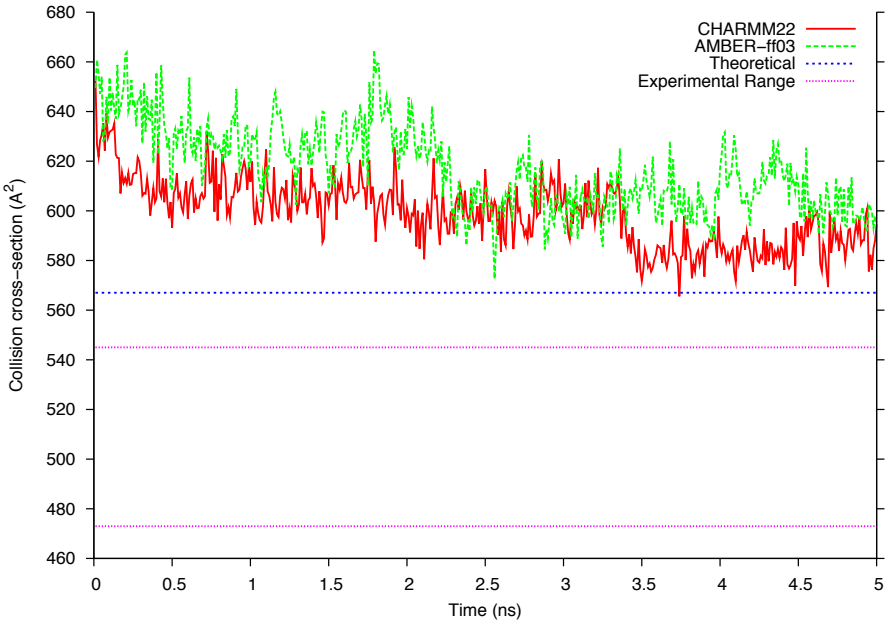


Figure 6.24: Collision cross-sections of the inter-sheet trimer simulations as a function of time, compared to the previously reported values of $[3M+2H]^{2+}$ obtained experimentally and via simulation.

spectively). While this was also applicable for the tetrameric system of TTR (as the initial predicted model consisted of 4 stacked layers of molecules), another possible arrangement of

TTR molecules was available. This third arrangement consisted of a dimer from one sheet and a dimer from a neighbouring sheet, henceforth referred to as the 'cuboid tetramer'.

6.4.4.1 Cuboid tetramer

Under the CHARMM22 parameterisation, the secondary structure of each residue over the course of the 5ns simulation (Fig. 6.25) revealed strong evidence of β-sheet character (intermolecular interactions) between the Ile(1)$_{107}$ & Pro(2)$_{113}$ residues (a feature observed over the initial 2ns of the intra-sheet trimer simulation). These two peptides were found to present stable 3_{10}- and α-helical content over Ala(1)$_{108}$-Leu(1)$_{110}$ and Ile(2)$_{107}$-Leu(2)$_{111}$ respectively. While the analysis indicated transient intermolecular β-sheet character between the Leu(3)$_{110}$ and Ala(4)$_{108}$ residues, the STRIDE algorithm did not reveal any indications of β-sheet like character between the two dimeric systems themselves. Visualisation of the tetramer system did suggest intermolecular backbone hydrogen bonding between the C-terminal of peptide 1, and the N-terminal of peptide 3 (Fig. 6.26a). Furthermore, the predominantly turn-like peptides 3 and 4 were found to remain extended and analysis of the average end-to-end distance over the last 1ns for these two peptides revealed both peptides to be defined by sharp peaks centred around 24.6Å.

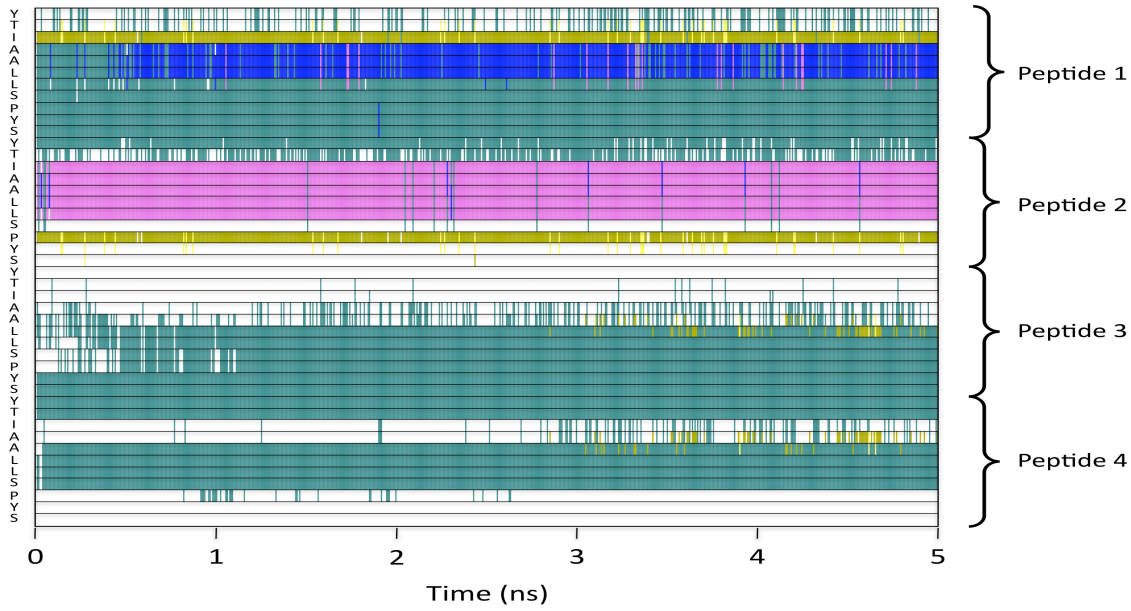


Figure 6.25: Secondary structure of each residue of the cuboid tetrameric TTR$_{105-115}$ under the CHARMM22 parameterisation as a function of time. Green corresponds to turn, white is unordered, pink is α-helix, blue is 3_{10}-helix and gold is β-strand.

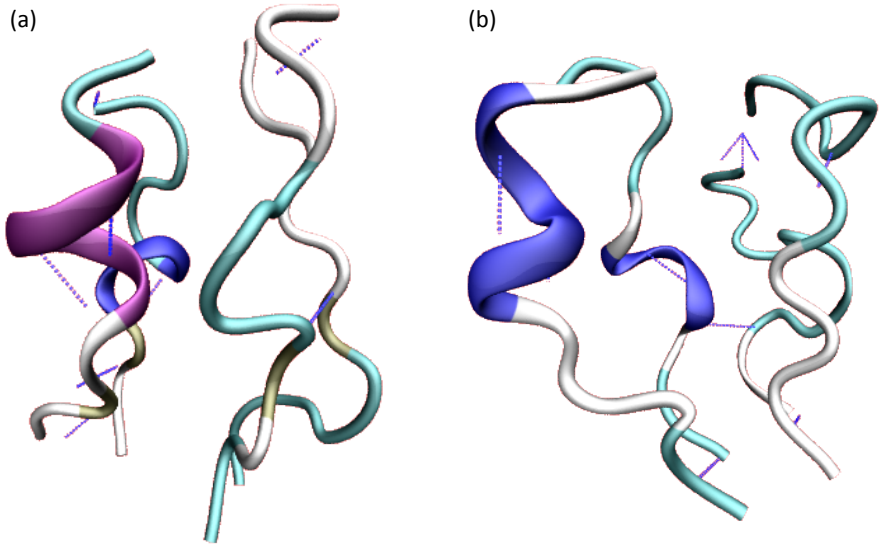


Figure 6.26: Visualisation of the structure of the $TTR_{105-115}$ cuboid tetramer after 5ns, under the (a) CHARMM22 and (b) AMBER-ff03 parameterisation. In relation to the initial structure, the peptide backbones were found to present RMSD values of 5.269Å and 6.319Å respectively.

Analysis of the secondary structure of each residue over time from the AMBER-ff03 system revealed that over the course of the initial 2.5ns of the simulation, the $Ala(1)_{109}$ and $Leu(2)_{111}$ residues possessed transient β-strand character (Fig. 6.27). This was coupled with a stable 3_{10}-helical motif spanning the $Ala(2)_{108}$-$Leu(2)_{110}$ residues over the course of the entire 5ns simulation; 3_{10}-helical character was also occasionally assigned to the $Ala(1)_{109}$-$Leu(1)_{111}$ residues. Furthermore the 3_{10}-helical motif spanning the $Ile(1)_{107}$-$Leu(1)_{110}$ residues, as captured in the visualisation of the system (Fig. 6.26b), was not found to be a dominant motif; the 'Peptide 1' was instead found to exhibit a flexible conformational ensemble over the later half of the simulation, consisting of a mixture of turn, unordered and aforementioned 3_{10}-helical structure. As under the CHARMM22 parameterisation, STRIDE did not reveal any evidence of intermolecular interactions between the two dimer systems. The other two peptides (denoted 'Peptide 3' and 'Peptide 4') were found to exhibit predominantly turn-like motifs, which were found to remain fairly extended (Fig. 6.26b).

In comparison to the intra-sheet dimeric systems, the simulations of the 'cuboid' tetramer under both parameterisations revealed some similarities in the predicted secondary structure of 'Peptide 1' and 'Peptide 2'. Under the CHARMM22 force field, both systems predicted helical motifs over the $Ala(1)_{108}$-$Leu(1)_{110}$ and $Ile(2)_{107}$-$Leu(2)_{111}$ regions; the latter was predomin-

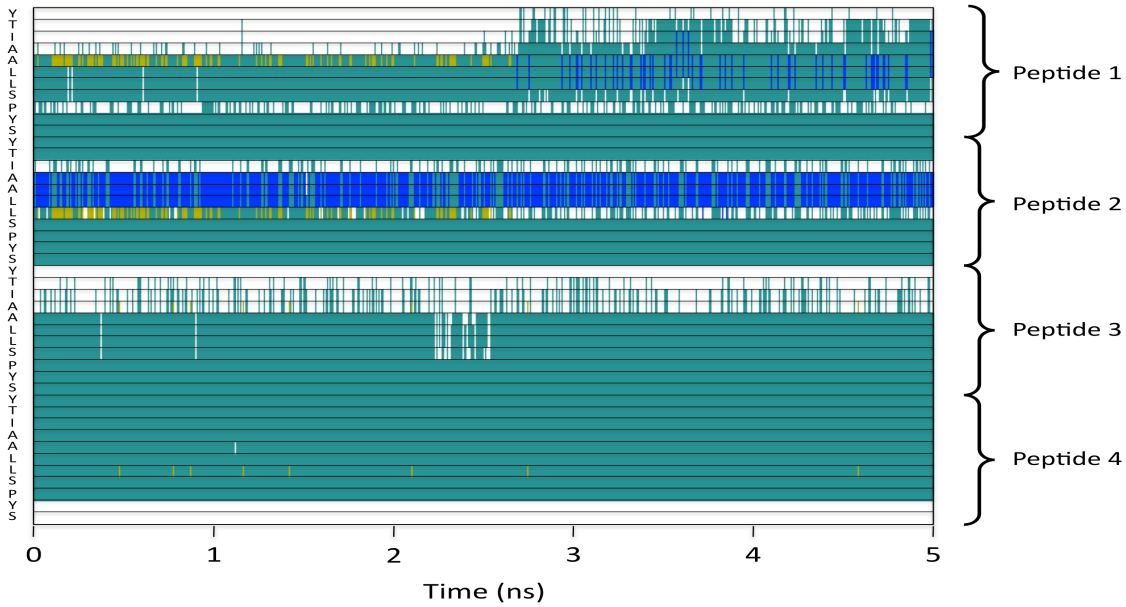


Figure 6.27: Secondary structure of each residue of the cuboid tetrameric $TTR_{105-115}$ under the AMBER-ff03 parameterisation as a function of time. Green corresponds to turn, white is unordered, pink is α-helix, blue is 3_{10}-helix and gold is β-strand.

antly α-helical in both the intra-sheet dimer and cuboid tetramer, while the former was a mixture of α- and 3_{10}-helix. The AMBER-ff03 parameterisation also revealed comparable 3_{10}-helical structure over the $Ala(2)_{108}$-$Leu(2)_{110}$ sequence, however while their was some evidence of helical structure over the $Ile(1)_{107}$-$Leu(1)_{111}$, such a motif was not as dominant in the cuboid tetramer as it was in the intra-sheet dimer.

6.4.4.2 Intra-sheet tetramer

Visualisation of the systems revealed that the 4 peptides remained fairly extended throughout the course of the simulation and were still relatively antiparallel with respect to one another (Fig. 6.28). The peptides appeared to have developed a slight twist under the CHARMM22 parameterisation, which was supported by analysis of the angles between the end-to-end vectors: 127.5° (∠**1.2**), 166.1° (∠**2.3**), 176.6° (∠**1.4**), while the angle between the two outermost peptides was 136.6° (∠**3.4**). In comparison, the AMBER-ff03 system was found to remain more aligned, with end-to-end vector angles of 172.8° (∠**1.2**), 172.5° (∠**2.3**) and 179.4° (∠**1.4**). After 5ns, initial β-sheet character between the two central peptides remained intact under the AMBER-ff03 force field. This suggested that the stability of β-sheet character may increase as a factor of

the size of intra-sheet oligomer.

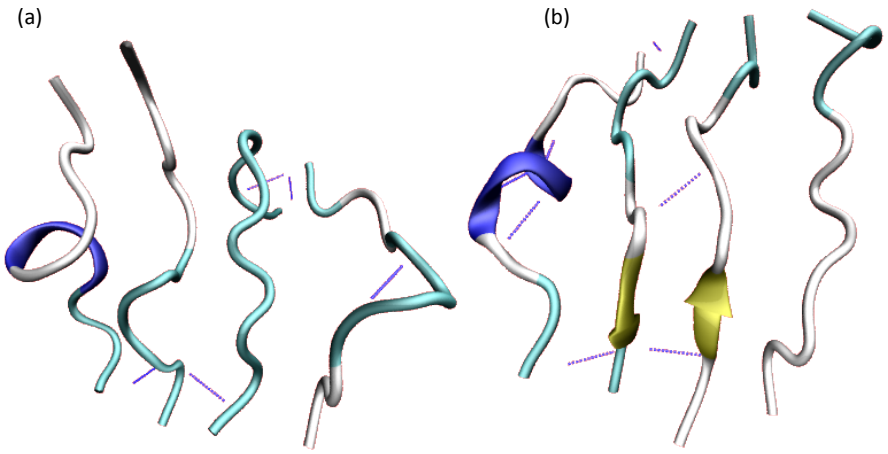


Figure 6.28: Visualisation of the structure of the $TTR_{105-115}$ intra-sheet tetramer after 5ns, under the (a) CHARMM22 and (b) AMBER-ff03 parameterisation. In relation to the initial structure, the peptide backbones were found to present RMSD values of 12.907Å and 13.699Å respectively.

Analysis of the secondary structure of each residue under the CHARMM22 parameterisation revealed that the observed 3_{10}-helical structure in one of the peptides was not a stable motif, but was only present as a transient feature (Fig. 6.29). The TTR(4) molecule was defined by an ensemble of turn and helical motifs over the 5ns; α- and 3_{10}-helical content was intermittently assigned to the Ala(4)$_{108}$-Leu(4)$_{111}$ and Ala(4)$_{108}$-Leu(4)$_{110}$ sequences respectively. Over the initial 1ns of the simulation, the STRIDE algorithm predicted β-sheet-like intermolecular interactions between the Ile(1)$_{107}$-Pro(2)$_{113}$, Leu(2)$_{110}$-Leu(3)$_{110}$ and Leu(1)$_{110}$-Ala(2)$_{108}$ pairings, with the latter assigned such character intermittently over the course of the 5ns simulation.

The simulation under the AMBER-ff03 force field revealed strong evidence of intermolecular interactions between the two central peptides, as evident through the Ile(1)$_{107}$-Ala(1)$_{108}$ and Ser(2)$_{112}$-Pro(2)$_{113}$ sequences being assigned β-sheet character throughout the course of the simulation. This was comparable to the evidence of β-sheet characteristics between the Ala(1)$_{108}$ and Ser(2)$_{112}$ residues in the intra-sheet trimer under the same parameterisation (Fig. 6.19) - a feature present over the course of the initial 3.5ns of that simulation. Furthermore, initial β-sheet character was also evident over the Leu(1)$_{111}$-Pro(1)$_{113}$ and Ile(2)$_{107}$-Ala(2)$_{109}$ residues, however this was lost after 0.5ns. The STRIDE secondary structure algorithm also revealed that the 3_{10}-helical structure evident in the visualisation of the tetramer (Fig. 6.28b) was not a dominant

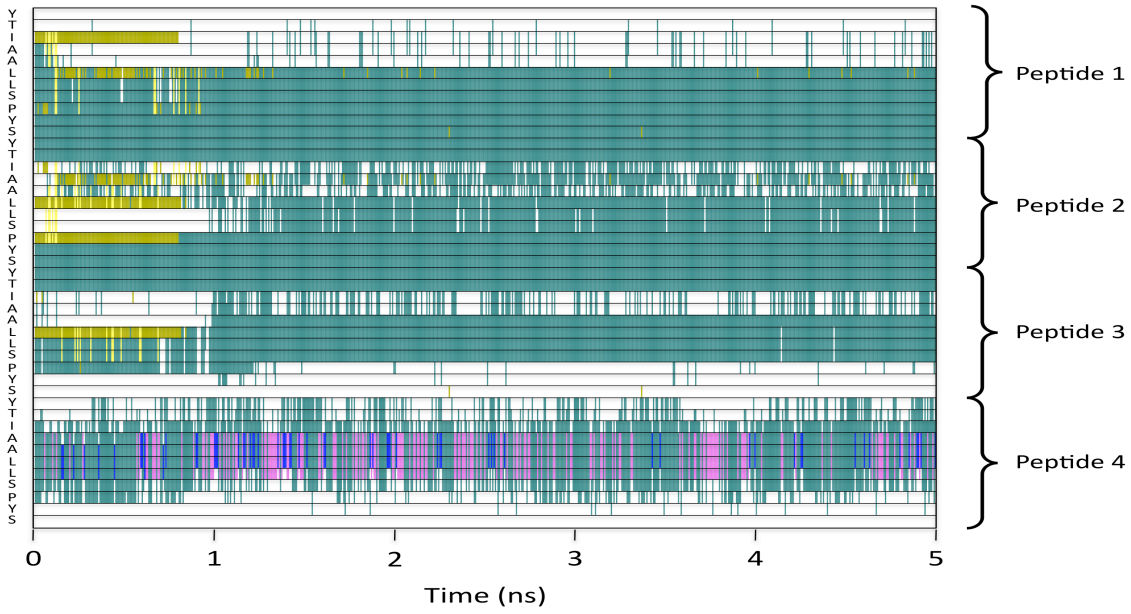


Figure 6.29: Secondary structure of each residue of the intra-sheet tetrameric $TTR_{105-115}$ under the CHARMM22 parameterisation as a function of time. Green corresponds to turn, white is unordered, pink is α-helix, blue is 3_{10}-helix, gold is β-strand and yellow is β-sheet.

structural motif over the course of the simulation, with the $Ala(3)_{108}$-$Leu(3)_{111}$ residues only assigned such structure during the last 200ps.

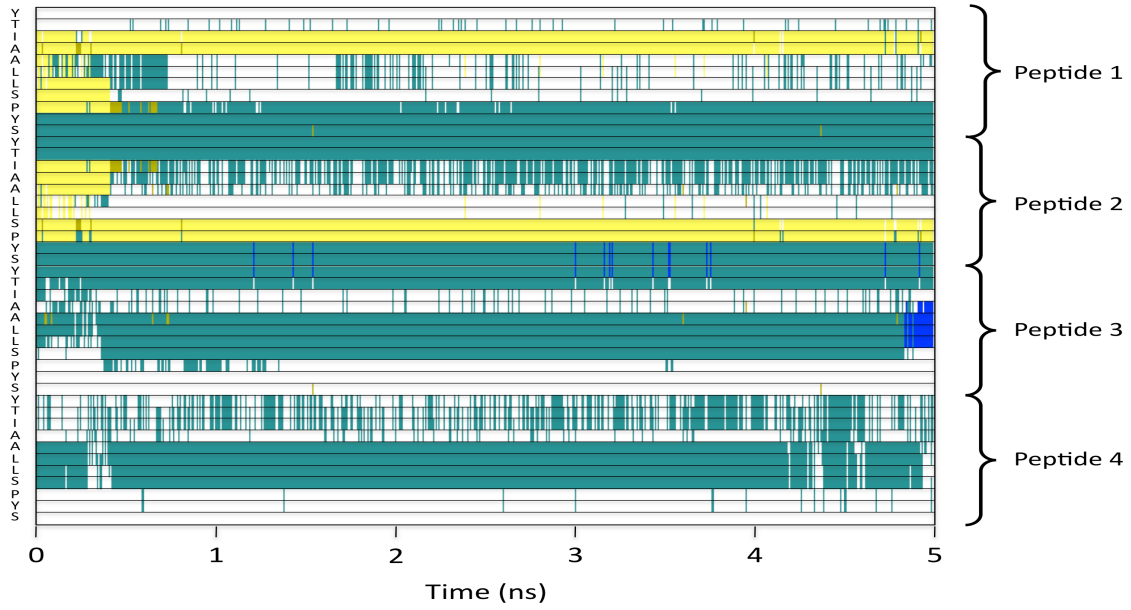


Figure 6.30: Secondary structure of each residue of the intra-sheet tetrameric $TTR_{105-115}$ under the AMBER-ff03 parameterisation as a function of time. Green corresponds to turn, white is unordered, pink is α-helix, blue is 3_{10}-helix, gold is β-strand and yellow is β-sheet.

In comparing the intra-sheet trimer and tetramer, interestingly the AMBER-ff03 force field appeared to reveal enhanced stabilisation of β-sheet structure in progressing from a trimeric system to a tetramer, whereas the CHARMM22 parameterisation showed a notable decrease in assigned β-sheet characteristics between the two simulations. In relation to this, analysis of the end-to-end distances revealed that the peptides under the AMBER-ff03 parameterisation were significantly more extended than their CHARMM22 counterparts; the average peptide lengths were 25.9Å (AMBER-ff03) and 20.675Å (CHARMM22). The peptides in the AMBER-ff03 intra-sheet tetramer were therefore found to be significantly more extended than in the intra-sheet trimer which was attributed to the increase in β-sheet character; no significant increase was detected for the CHARMM22 systems. It should be noted however that both simulations predicted the peptides in the intra-sheet tetramer to remain fairly aligned with one another, revealing that there were significant intermolecular interactions in order to retain such a geometry.

6.4.4.3 Inter-sheet tetramer

Visualisation of the two inter-sheet tetramers revealed a significant difference in the predicted structure of the system (Fig. 6.31). While the peptides under the CHARMM22 parameterisation remained aligned with respect to one another displaying a slight twist (angles between neighbouring peptide end-to-end vectors were found to range from 154.7° to 174.8°), the tetramer subjected to the AMBER-ff03 force field resulted in two coiled dimers perpendicular to one another: 175.1° (∠**1.3**), 164.8° (∠**2.4**), 62.9° (∠**1.2**). Such a result was unexpected, as this was the only system to present such a dramatic difference in the tertiary structure from utilising different force fields. However in comparison to the simulations of the charged species conducted previously [209], there did appear to be similarities in the structure of the tetramer system reported by the authors simulated annealing algorithm.

Over the course of the 5ns simulation, the CHARMM22 system was found to be comprised of three stable helical structures spanning the Ile(1)$_{107}$-Leu(1)$_{111}$, Tyr(4)$_{105}$-Ala(4)$_{109}$ and Thr(3)$_{106}$-Leu(3)$_{111}$ sequences consisting of a mixture of α- 3_{10}-helical motifs; the Thr(3)$_{106}$-Leu(3)$_{111}$ helix was found to initiate from a 3_{10}-like motif spanning Ala(3)$_{109}$-Leu(3)$_{111}$, expanding after 2.5ns. While the STRIDE secondary structure assignment algorithm assigned turn-like motifs to the TTR molecule designated 'Peptide 2', visualisation of the structure of

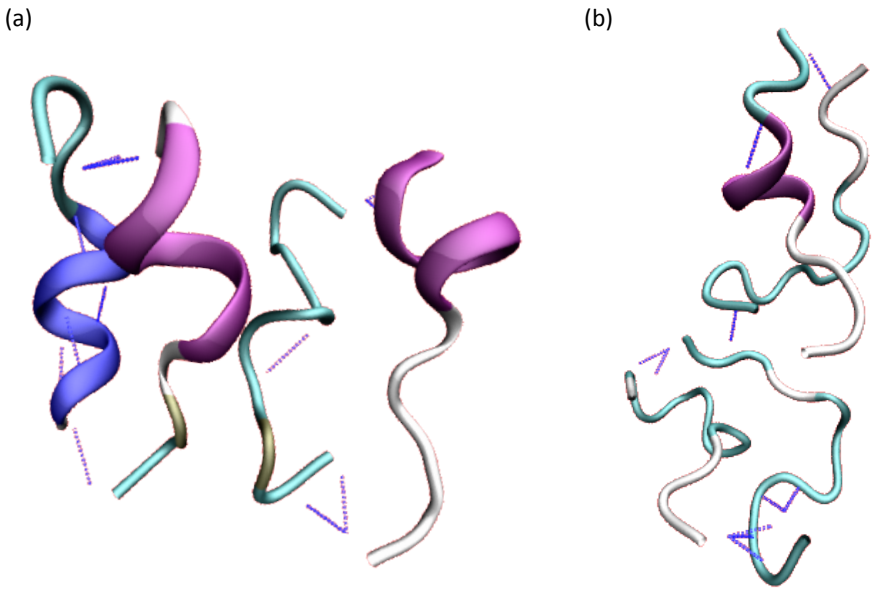


Figure 6.31: Visualisation of the structure of the $TTR_{105-115}$ inter-sheet tetramer after 5ns, under the (a) CHARMM22 and (b) AMBER-ff03 parameterisation. In relation to the initial structure, the peptide backbones were found to present RMSD values of 7.437Å and 12.086Å respectively.

the tetramer revealed this peptide to remain fairly extended. Furthermore, this peptide was found to form stable intermolecular interactions with its neighbour 'Peptide 1' as evident through the assignment of β-sheet character between $Pro(1)_{113}$ and $Ile(2)_{107}$.

β-sheet character was also evident under the AMBER-ff03 parameterisation, with STRIDE assigning such structure between the $Leu(2)_{110}$ and $Ser(4)_{112}$ residues. The collapse of this structure was found to coincide with the formation of a helical motif after 3ns: initially a 3_{10}-helix before forming an α-helix spanning $Ala(4)_{108}$-$Leu(4)_{111}$. The secondary structure of the other three peptides was predominantly characterised as unordered and turn-like.

The formation of helical structure under the CHARMM22 parameterisation was not unexpected. Both the inter-sheet dimer and trimer revealed that without the presence of initial β-sheet structure, the TTR molecules displayed significant helical propensity (Fig. 6.14, 6.22). However the presence of stable β-sheet characteristics between two peptides in both the CHARM22 and AMBER-ff03 systems was an unforeseen feature, as the none of the previous simulations involving strands from differing layers had revealed such structuring.

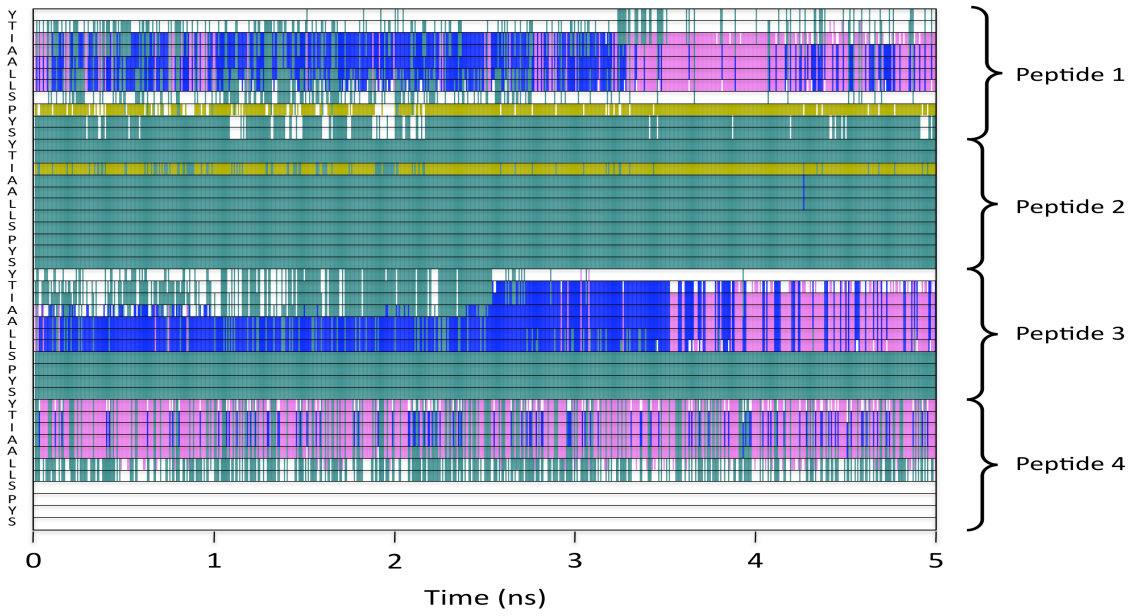


Figure 6.32: Secondary structure of each residue of the inter-sheet tetrameric $TTR_{105-115}$ under the CHARMM parameterisation as a function of time. Green corresponds to turn, white is unordered, pink is α-helix, blue is 3_{10}-helix and gold is β-strand.

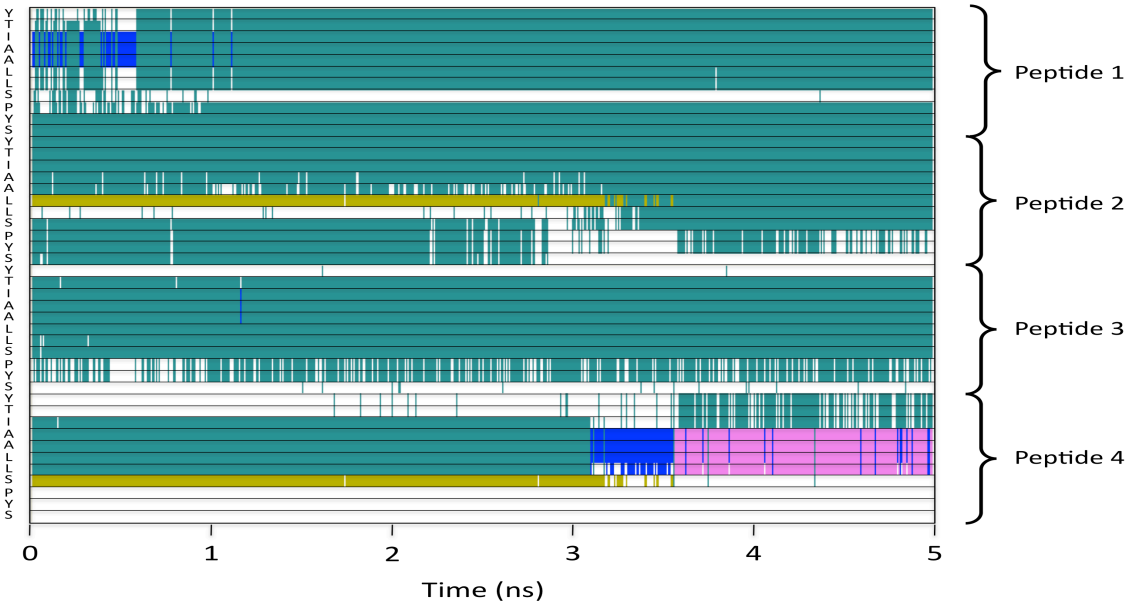


Figure 6.33: Secondary structure of each residue of the inter-sheet tetrameric $TTR_{105-115}$ under the AMBER-ff03 parameterisation as a function of time. Green corresponds to turn, white is unordered, pink is α-helix, blue is 3_{10}-helix and gold is β-strand.

6.4.4.4 Tetramer Cross-sections

The collision cross-sections were evaluated for all three tetramer systems under the CHARMM22 (Fig. 6.34) and AMBER-ff03 (Fig. 6.35) parameterisations. Interestingly, under both para-

meterisations, the 'cuboid' tetramer was found to adopt a more compact geometry than the previously reported simulations; over the last 1ns, AMBER-ff03 resulted in an average collision cross-section of 708Å^2, CHARMM22 731Å^2, *cf.* 737Å^2 for the previously simulated [4M+3H]$^{3+}$ [209]. The values calculated herein were found to differ from the experimental value of 663Å^2 (also for a [4M+3H]$^{3+}$ system) by 6.79% and 10.26% respectively and presented the second smallest deviation from the experimental mean; the smallest deviation was found in the monomer system (5.47%, AMBER-ff03 and 6.64%, CHARMM22).

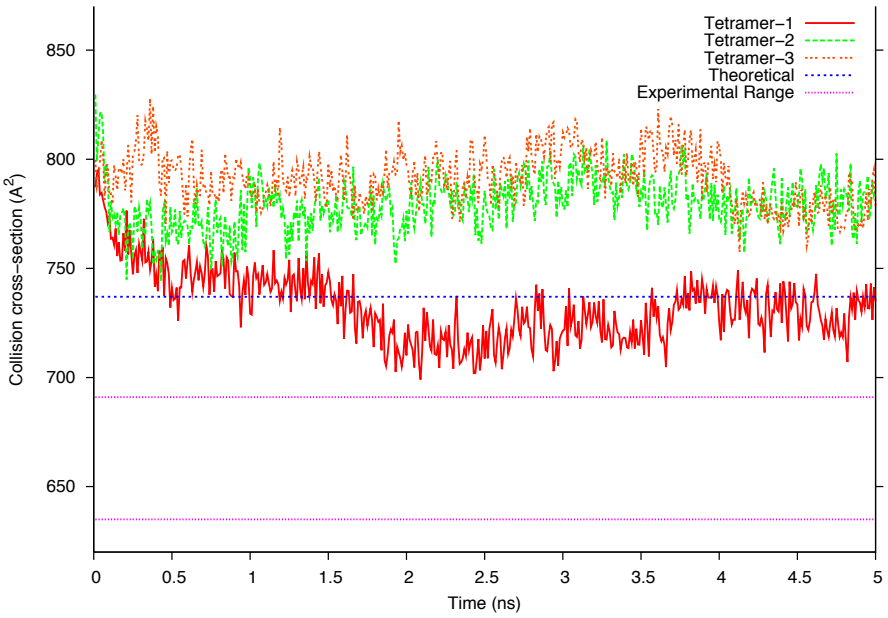


Figure 6.34: Collision cross-sections of the tetramer simulations under the CHARMM22 parameterisation as a function of time, compared to the previously reported values of [4M+3H]$^{3+}$ obtained experimentally and via simulation. Tetramer 1, 2, and 3, refer to 'cuboid', 'intra', and 'inter' tetramers respectively.

Such a sudden decrease in deviation from experimental mean was unexpected, and if the neutral species in any way reflected the actual species measured experimentally, then the cuboid tetramer could be an unrealistic geometry. In comparison, the intra-sheet tetramer was found to result in collisional cross-sections of 779Å^2 and 767Å^2 for the CHARMM22 and AMBER-ff03 systems respectively. These results equated to deviations from the previous simulations of 5.70% (CHARMM22) and 4.07% (AMBER-ff03) and the experimental mean by 17.5% (CHARMM22) and 15.7% (AMBER-ff03). These values were found to be comparable to those obtained in the dimeric and trimeric systems, thereby bringing the validity of the cuboid tetramer arrangement

into question.

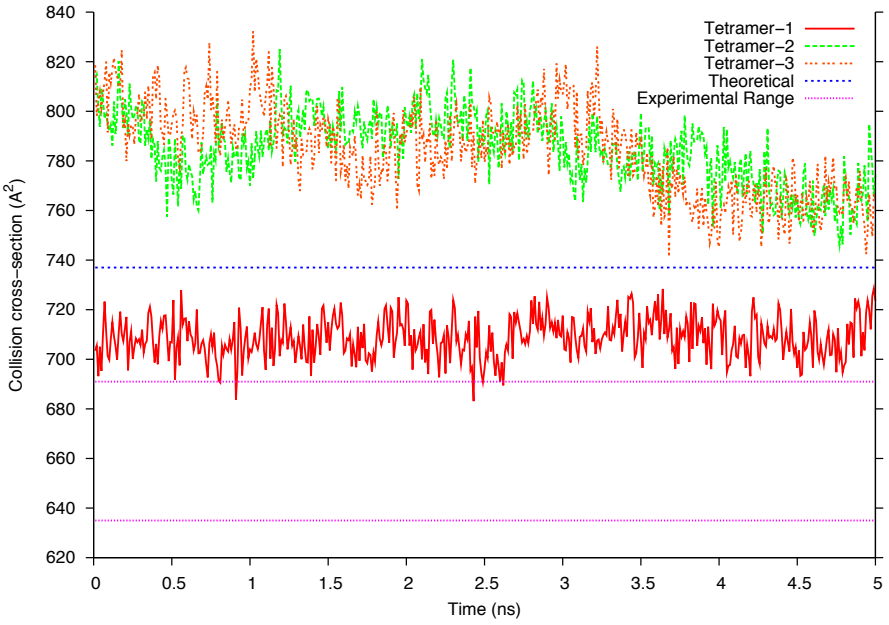


Figure 6.35: Collision cross-sections of the tetramer simulations under the AMBER-ff03 parameterisation as a function of time, compared to the previously reported values of $[4M+3\mathrm{H}]^{3+}$ obtained experimentally and via simulation. Tetramer 1, 2, and 3, refer to 'cuboid', 'intra', and 'inter' tetramers respectively.

The inter-sheet tetramer similarly resulted in collisional cross-sections of 778Å^2 and 763Å^2 under the CHARMM22 and AMBER-ff03 parameterisations, which corresponded to deviations from the previous simulations of 5.56% (CHARMM22) and 3.53% (AMBER-ff03) and the previous experiments by 17.3% (CHARMM22) and 15.1% (AMBER-ff03).

6.4.5 Comparison to Experiment

The previous investigation into the early aggregates of $\mathrm{TTR}_{105-115}$ revealed that the collision cross-sections of the oligomers increased linearly with respect to the highest charge state of each oligomer [209]. The authors noted that oligomers above a trimer resulted in near identical cross-sections for identical charges (i.e. tetramer and pentamer of charge 3+). The simulations presented herein resulted in a linear increase for both an intra- and inter-sheet progression under both parameterisations (Fig. 6.36) which was inline with the observed trend in the oligomers up to the tetramer; it should be noted that the reported species did increase in charge and the linear relationship for the highest charge state in the full set of oligomers was rationalised as

due to a competition between packing and Coulombic repulsion. As the species studied above should not have presented such a repulsion, then logically one would have expected the collision cross-sections to be on par or smaller than those reported previously. As evident in the values for the neutral species, the rate of increase was found to be higher for the neutral species than the highest charge state oligomers (Fig. 6.36). Furthermore, as the neutral simulations extended to only a tetrameric system, no comparison could be made in regards to the observed trend in the oligomers of 4 or more peptides and of identical charges.

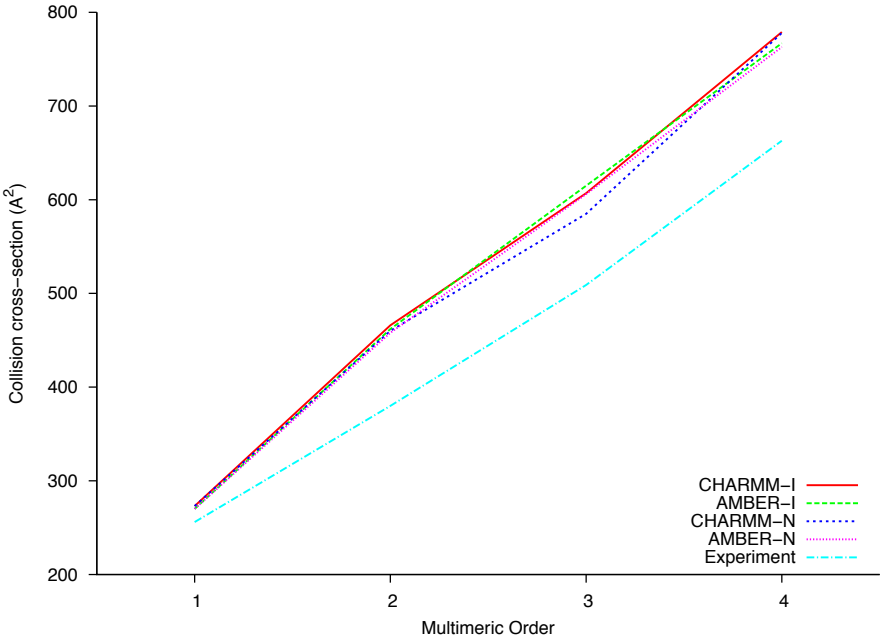


Figure 6.36: Observed collision cross-sections of simulations over last 1-2ns as a function of size of oligomer under the CHARMM22 and AMBER-ff03 parameterisation. Series are divided into extensions from within a sheet (Intra-sheet, 'I') or between sheets (Inter-sheet, 'N'). Values are compared to the experimentally obtained collision cross-sections.

6.4.6 Comparison to Simulations in Solution

In addition to simulations conducted on charged oligomers, previous investigation of $TTR_{105-115}$ has focused on utilising molecular dynamics to explore the process of self-assembly and the resulting structures of aggregates of neutral $TTR_{105-115}$ peptides in solution (using an implicit solvent model) [208], focusing on the monomeric, trimeric and tetrameric aggregates: dimeric aggregates have been investigated separately [210]. The monomeric species was found to exhibit a high propensity for α-helical content between the Thr_{106}-Leu_{110} residues, which the authors

noted was consistent with experimental evidence of $TTR_{105-115}$ displaying propensity for α-helical or β-turn content [212]; experimentally the peptide was however found to remain predominantly unordered. While the simulations of the monomer species conducted herein did not display any α-helical content under either parameterisation, they did reveal evidence of turn-like structure in agreement with the previous experimental findings [212].

The propensity for helical content was however observed in the dimeric, trimeric and tetrameric simulations conducted as part of this body of work; such structure was observed particularly in the inter-sheet oligomers. This would suggest that as the inter-sheet models resulted in similar structure to the monomer in solution, one could regard each peptide in these systems as an isolated chain. Furthermore, while the α-helical propensity was found to decrease with respect to size of oligomer, the authors noted that the propensity for β-strand conformation increased. This was certainly consistent with the observed trends in the intra-sheet models conducted herein, which revealed the AMBER-ff03 parameterisation to display increased stability in β-sheet structure.

Having revealed similarities in the secondary structural trends in the oligomers and differences in the reported end-to-end distances for the monomeric species, it was of interest to compare the end-to-end distances of the gas phase and solution trimer simulations. To that end, the distance between the N atom of the Tyr_{105} N-terminus and the C atom of the Ser_{115} C-terminus were evaluated for every peptide over the last 1ns in each of the simulations conducted herein (Tab. 6.1). The original starting structure was characterised by a distance of $\sim$35.7Å, therefore the length of all peptides in the gas phase would appear to have contracted by at least 8.3Å (based on the maximum length of 27.4Å). The authors of the investigation into the solution structure of $TTR_{105-115}$ aggregates defined a peptide to be extended of the average end-to-end distance was in excess of 20Å, a value chosen as it was between than the average distance in the monomeric simulation and the fully extended length (one can regard this as similar to that of the initial starting length of 35.7Å); within the crystal structure of native $TTR_{105-115}$, this distance is found to be 22.6Å (Protein Data Bank ID code 1BMZ) [213].

Of the four trimer systems studied herein, only the CHARMM22 intra-sheet trimer was found to result in an average length in excess of the 20Å cut-off for an extended conformation. Furthermore, all trimer systems were found to be more contracted than the average lengths of 25Å re-

Force Field	Monomer	Dimer		Trimer		Tetramer		
		Intra	Inter	Intra	Inter	Cuboid	Intra	Inter
CHARMM22	3.14	13.8	14.2	20.6	12.2	19.2	22.2	16.6
		16.2	14.0	21.8	11.3	19.8	20.5	15.0
				18.6	16.6	24.6	20.2	14.6
						24.6	19.8	19.8
AMBER-ff03	2.95	13.8	14.2	19.4	18.2	19.4	27.4	14.6
		17.8	17.8	17.0	16.6	18.2	25.8	17.0
				17.4	21.0	19.8	24.6	11.4
						14.2	25.8	18.6

Table 6.1: Separation between N atom of the Tyr_{105} N-terminus and C atom of the Ser_{115} C-terminus over the last 1ns in each simulation. The values are representative of each peptide in the oligomer (e.g. Peptide 1, Peptide 2, etc.) and are all in Å.

ported for trimer systems in solution; the length of each peptide in the $TTR_{105-115}$ trimer in solution was found to vary between 13Å and 32Å, and only the CHARMM22 inter-sheet trimer resulted in lengths outwith this range. From the results of the simulations, the intra-sheet systems generally resulted in more extended conformations as the size of oligomer increased, which was coupled with the increase in stability in β-sheet conformation. This was exemplified in the intra-sheet tetramer under the AMBER-ff03 parameterisation, which was regarded as displaying the greatest β-sheet content. This system was found to result in the most extended end-to-end distances of all the models tested (Tab. 6.1); it should be noted that there was obviously some contraction from the original starting length of $\sim$35.7Å, however the distances recorded were in excess of that in the native TTR crystal structure.

6.5 Conclusions

Intermolecular interactions between the $TTR_{105-115}$ molecules were apparent from the dimeric to tetrameric systems, with the AMBER-ff03 force field presenting the greater amount of such structure; only the intra-sheet trimer was found to display any significant β-sheet characteristics under CHARMM22, while the intra-sheet tetramer and dimer displayed little and none respectively. Such structure was only apparent in the intra-sheet dimer and trimer models, while all tetramer models presented evidence of β characteristics under both parameterisations. It can be concluded that such assignment may be more apparent and stable (especially under AMBER-

ff03) as the size of the oligomer increases. Helical structure may persist in the inter-sheet systems due to the apparent high affinity for such motifs, and as evident from the tetramer system, β-sheet structuring may also become more apparent and could potentially supersede that of the helix. In the simulations studied herein, both intra- and inter-sheet systems resulted in comparable collision cross-sections. This may present a difficulty in establishing which of the two configurations was measured experimentally. The possibility of extended cuboid-esque starting structures may also reveal whether such an initial geometry consistently results in smaller collisional cross-sections than their intra- and inter-sheet counterparts. The collision cross-sections of all neutral species were found to be higher than anticipated; the increase in size in the charge series was attributed to coulombic repulsion, which would be absent in the neutral species. While the recorded collision cross-sections of certain models appeared to be steady over the course of the 5ns simulations, all species were shown to be significantly shorter than their original extension and further contraction could not be ruled out. As such, the simulation times may be too short in order to draw meaningful comparisons with the experimentally recorded collision cross-sections. In regards to the previous simulation, while the conventional MD simulations were of comparable length (both 5ns), the reported collision cross-sections were from subsequent simulated annealing (not performed in this study) [209]. As the study revealed the peptide geometries to be more collapsed when subjected to simulated annealing, it is of note that the simulations conducted herein only resulted in collision cross-sections differing by only as much as 8.5% than those reported previously.

7

Conclusions and Future Outlook

The aims of this doctoral study have centred around investigating whether small model peptide systems provided insight into biological phenomena such as antibody discrimination, viral fusion or self-assembly, and whether the MD simulations reflected results obtained experimentally. From the peptides studied herein, one can conclude that the simulations have had some success in providing information about the systems of interest. Furthermore, the peptide systems studied within were found to be characterised by rugged free-energy landscapes unlike their protein counterparts (defined by singular, deep minima). These landscapes were subsequently shown to be highly plastic and sensitive to changes in the local environment.

In the investigation into the glycoprotein gp41 covered in *Chapters 3 & 4*, the models have revealed secondary structure propensity supported by utilising circular dichroism. Furthermore, these models have predicted tryptophan residues to be contained within a hydrophobic core; fluorescence spectroscopy suggested that these residues would not be exposed in order to interact with a membrane. The systems representing antibody epitopes investigated in *Chapter 5*, have suggested that such models may be of significant use in rationalising differences in antibody-antigen interactions. These peptides have indicated that variations in binding affinities may be attributed to minute alterations in the peptide backbone and that MD provides the high sensitivity required for evaluating such structure. Finally the investigation into the formation of amyloid aggregations in *Chapter 6* has highlighted that in utilising models systems, care should be taken in regards to the choice of force field used to represent the system; force fields have been shown

to differ widely in the regions of dihedral space sampled, as suggested in a previous in depth study [32].

While conclusions have been drawn on the basis of work performed for this thesis, a number of projects could be taken up in order to ensure a thorough investigation into the areas covered in this work in addition to advancing our current understanding of the molecules of interest.

7.1 GP41

Having simulated the monomeric gp41$_{532-544}$ and gp41$_{659-671}$, in addition to the dimer, the next logical step would be to simulate these systems in the presence of membranes. This would reveal whether the reported interplay was measurable by means of molecular dynamics and if the gp41$_{532-544}$ peptide exhibited β-sheet structure in the presence of anionic membranes. The experimentally observed stabilisation of helical content in the gp41$_{532-544:659-671}$ dimer may be difficult to elucidate using the current molecular dynamics methodology, as the simulation under aqueous conditions displayed a high degree of helical content. In order to compensate for this, it may be necessary to explore additional force fields or water models, for example the explicit 4-site TIP4P force field or implicit water models; however, some implicit models have been shown to overestimate the degree of secondary structure and number of intramolecular hydrogen bonds [38], and therefore may not be appropriate for this system.

7.2 hCG

Having revealed minute differences in the structure of hCGβ_{66-80} and LHβ_{86-100}, it would be natural to extend this investigation to include the corresponding full proteins. This would reveal whether the variations detected in the peptide fragments were retained in the native structure of the glycoproteins human chorionic gonadotropin and luteinizing hormone respectively. The aim of which would be to explore if studying model peptide fragments representative of antibody epitopes resulted in meaningful structural information and helped to rationalise antibody discrimination; certainly, the study conducted herein was found to support the findings of point mutation binding results by suggesting that hCGβ_{66-80} and the histidine mutant shared structural features lacking in the other three main models tested. If so, such an approach could be applied

to a theoretically unlimited number of antibody-antigen systems. Furthermore, the synergistic approach applied in this study is unashamedly believed to be the correct strategy in exploring such biomolecular phenomena, and could lead to high-value commercial products.

7.3 TTR

The simulations of $TTR_{105-115}$ investigated herein have revealed β-sheet structure to be stabilised (from an intra-sheet perspective) as the size of the oligomer was increased. It would therefore be interesting to investigate these peptide fragments in the presence of water to probe the competition of intermolecular interactions between peptide & peptide and peptide & water, revealing whether the presence of water had an impact on the stability of β-sheet conformation as a factor of oligomer size. Previous MD investigations into $TTR_{105-115}$ have focused on the aggregation utilising an implicit solvent model [208, 210]. As such solvents may overestimate the number of intramolecular hydrogen bonds, it may be of interest to investigate the stability of $TTR_{105-115}$ oligomers utilising an explicit water model such as TIP3P (used throughout this doctoral study).

References

[1] R. W. Woody. *Theory of Circular Dichroism of Proteins*. In G. D. Fasman, editor, *Circular Dichroism and the Conformational Analysis of Biomolecules*. Springer, 1996.

[2] G. Barbato, E. Bianchi, P. Ingallinella, W. H. Hurni, M. D. Miller, G. Ciliberto, R. Cortese, R. Bazzo, J. W. Shiver, and A. Pessi. *Structural Analysis of the Epitope of the Anti-HIV Antibody 2F5 Sheds Light into Its Mechanism of Neutralization and HIV Fusion*. *J. Mol. Biol.*, 330:1101–1115, 2003.

[3] V. Buzon, G. Natrajan, D. Schibli, F. Campelo, M. M. Kozlov, and W. Weissenhorn. *Crystal Structure of HIV-1 gp41 Including Both Fusion Peptide and Membrane Proximal External Regions*. *PLoS Pathog.*, 6:e1000880, 2010.

[4] A. J. Lapthorn, D. C. Harris, A. Littlejohn, J. W. Lustbader, R. E. Canfield, K. J. Machin, F. J. Morgan, and N. W. Isaacs. *Crystal structure of human chorionic gonadotropin*. *Nature*, 369:455–461, 1994.

[5] C. E. MacPhee. *Theoretical model of TTR$_{105-115}$ antiparallel β-sheet stacking*. Unpublished data.

[6] D. Hames and N. Hooper. *Biochemistry*. Taylor & Francis, 2005.

[7] S. Krimm and J. Bandekar. *Vibrational spectroscopy & conformation of peptides, polypeptides and proteins*. *Adv. Prot. Chem.*, 38:181–367, 1986.

[8] C.M. Venkatachalam. *Stereochemical criteria for polypeptides and proteins. V. Conformation of a system of three linked peptide units*. *Biopolymers*, 6:1425–1436, 1968.

[9] J.S. Richardson. *The anatomy and taxonomy of protein structure*. *Adv. Protein Chem.*, 34:167–339, 1981.

[10] E. G. Hutchinson and J. M. Thornton. *A revised set of potentials for β-turn formation in proteins*. *Protein Science*, 3:2207–2216, 1994.

[11] P. Y. Chou and G. D. Fasman. *Prediction of protein conformation*. *Biochemistry*, 13:222–245, 1974.

[12] G. M. Whitesides and M. Boncheva. *Beyond molecules: self-assembly of mesoscopic and macroscopic components*. *Proc. Natl. Acad. Sci.*, 99:4769–4774, 2002.

[13] G. M. Whitesides, J. P. Mathias, and C. T. Seto. *Molecular self-assembly and nanochemistry: a chemical strategy for the synthesis of nanostructures*. *Science*, 254:1312–1319, 1991.

[14] S. Zhang, D. M. Marini, W. Hwang, and S. Santoso. *Design of nanostructured biological materials through self-assembly of peptides and proteins*. *Current Opinion in Chemical Biology*, 6:865–871, 2002.

[15] K. E. Drexler. *Building molecular machine systems*. *Trends Biotechnol.*, 17:5–7, 1999.

[16] R. C. Merkle. *Biotechnology as a route to nanotechnology*. *Trends Biotechnol.*, 17:271–274, 1999.

[17] A. Aggeli, M. Bell, L. M. Carrick, C. W. G. Fishwick, R. Harding, P. J. Mawer, S. E. Radford, A. E. Strong, and N. Boden. *pH as a trigger of peptide β-sheet self-assembly and reversible switching between nematic and isotropic phases*. *J. Am. Chem. Soc.*, 125:9619–9628, 2003.

[18] K. Lu, J. Jacob, P. Thiyagarajan, V. P. Conticello, and D. G. Lynn. *Exploiting amyloid fibril lamination for nanotube self-assembly*. *J. Am. Chem. Soc.*, 125:6391–6393, 2003.

[19] S. Santoso, H. Hwang, H. Hartman, and S. Zhang. *Self-assembly of surfactant-like peptides with variable glycine tails to form nanotubes and nanovesicles*. *Nano Lett.*, 2:687–691, 2002.

[20] G. Xu, W. Wang, J. T. Groves, and M. H. Hecht. *Self-assembled monolayers from a designed combinatorial library of de novo β-sheet proteins*. *Proc. Natl. Acad. Sci.*, 98:3652–3657, 2001.

[21] S. K. Maji, D. Haldar, A. Banerjee, and A. Banerjee. *Fibril-forming model synthetic peptides containing 3-aminophenylacetic acid*. *Tetrahedron*, 58:8695–8702, 2002.

[22] A. Aggeli, M. Bell, N. Boden, J. N. Keen, P. F. Knowles, T. C. B. McLeish, M. Pitkeathly, and S. E. Radford. *Responsive gels former by the spontaneous self-assembly of peptides into polymeric β-sheet tapes*. *Nature*, 386:259–262, 1997.

[23] J. C. Rochet and P. T. Lansbury Jr. *Amyloid fibrillogenesis: themes and variations*. *Curr. Opin. Struct. Biol.*, 10:60–68, 2000.

[24] M. Goedert, M. G. Spillantini, and S. W. Davies. *Filamentous nerve cell inclusions in neurodegenerative diseases*. *Curr. Opin. Neurobiol.*, 8:619–632, 1998.

[25] S. B. Prusiner. *Prions*. *Proc. Natl. Acad. Sci.*, 95:13363–13383, 1998.

[26] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*. Clarendon Press, 1989.

[27] A. Leach. *Molecular Modelling: Principles and Applications*. Prentice Hall, 2001.

[28] D. Frenkel and B. Smit. *Understanding Molecular Simulation: From Algorithms to Applications*. Academic Press, 2001.

[29] B.R. Brooks, C. L. Brooks III, A. D. MacKerell Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Fieg, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus. *CHARMM: The Biomolecular Simulation Program*. *J. Comput. Chem.*, 30:1545–1614, 2009.

[30] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. *A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J. Am. Chem. Soc.*, 117:5179–5197, 1995.

[31] A. MacKerell, D. Bashford, M. Bellott, R. Dunbrack, J. Evanseck, M. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. Lau, C. Mattos, S. Michnick, T. Ngo, D. Nguyen, B. Prodhom, W. Reiher, B. Roux, M. Schlenkrich, J. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. *All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins*. *J. Phys. Chem. B*, 102:3586–3616, 1998.

[32] R. B. Best, N.-V. Buchete, and G. Hummer. *Are Current Molecular Dynamics Force Fields too Helical? Biophysical Journal*, 95:L07–L09, 2008.

[33] A. MacKerell. *Empirical Force Fields for Biological Macromolecules: Overview and Issues*. *J. Comput. Chem.*, 25:1584–1604, 2004.

[34] A. MacKerell, M. Feig, and C. L. Brooks III. *Improved Treatment of the Protein Backbone in Empirical Force Fields*. *J. Am. Chem. Soc.*, 126:698–699, 2004.

[35] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling. *Comparison of multiple Amber force fields and development of improved*. *Proteins*, 65:712–725, 2006.

[36] Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, and T. Lee. *A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations*. *J. Comput. Chem.*, 24:1999–2012, 2003.

[37] M.C. Lee and Y. Duan. *Distinguish protein decoys by using a scoring function based on a new Amber force field, short molecular dynamics simulations, and the generalized Born solvent model*. *Proteins*, 55:620–634, 2004.

[38] M. J. Feige and E. Paci. *Rate of Loop Formation in Peptides: A Simulation Study*. *J. Mol. Biol.*, 382:556–565, 2008.

[39] S. M. Kelly and N. C. Price. *The application of circular dichroism to studies of protein folding and unfolding*. *Biochim. Biophys. Acta.*, 1338:161–185, 1997.

[40] S. M. Kelly and N. C. Price. *The Use of Circular Dichroism in the Investigation of Protein Structure and Function*. *Curr. Protein Pept. Sci.*, 1:349–384, 2000.

[41] R. W. Woody. *Circular dichroism*. *Methods Enzymol*, 246:34–71, 1995.

[42] J. F. Towell III and M. C. Manning. *Analysis of protein structure by circular dichroism spectroscopy*. In N. Purdie and H. G. Brittain, editors, *Analytical Applications of Circular Dichroism*. Elsevier Science B. V., 1994.

[43] B. A. Wallace. *Synchrotron radiation circular-dichroism spectroscopy as a tool for investigating protein structures*. *J. Synchrotron Rad.*, 7:289–295, 2000.

[44] W. C. Johnson Jr. *Protein secondary structure and circular dichroism: A practical guide*. *Proteins Struct. Funct. Genet.*, 7:205–214, 1990.

[45] A. Perczel and G. D. Fasman. *Quantitative analysis of cyclic β-turn models*. *Protein Science*, 1:378–395, 1992.

[46] B.I. Baello, P. Pancoska, and T.A. Keiderling. *Enhanced prediction accuracy of protein secondary structure using hydrogen exchage fourier transform IR spectrscopy*. *Analytical Biochemistry*, 280:46–57, 2000.

[47] A.T. Tu. *Raman Spectroscopy in biology - Principles & Applications*. Wiley-Interscience, New York, 1982.

[48] T. Miyazawa, T. Shimanouchi, and S.I. Mizushima. *Normal vibrations of N-Methylacetamide*. *Journal of Chemical Physics*, 29:611–616, 1958.

[49] A. Barth and C. Zscherp. *What vibrations tell us about proteins*. *Quality reviews of Biophysics*, 35:369–430, 2002.

[50] R. Schweitzer-Stenner. *Advances in vibrational spectroscopy as a sensitive probe of peptide and protein structure. A critical review. Vibrational Spectroscopy*, 42:98–117, 2006.

[51] J. Bandekar. *Amide modes and protein conformation*. *Biochimica et Biophysica Acta*, 1120:123–143, 1991.

[52] P. I. Haris and D. Chapman. *The conformational Analysis of Peptides using Fourier Transform IR Spectroscopy*. *Biopolymers*, 37:251–263, 1995.

[53] M. Hollósi, Z.S. Majer, A.Z. Rónai, A. Magyar, K. Medzihradszky, S. Holly, A. Perczel, and G.D. Fasman. *CD & FTIR Spectroscopic studies of Peptides: II. Detection of β-turns in Linear Peptides*. *Biopolymers*, 34:177–185, 1994.

[54] W. K. Surewicz, H. H. Mantsch, and D. Chapman. *Determination of Protein Secondary Structure by Fourier Transform Infrared Spectroscopy: A Critical Assessment*. *Perspectives in Biochemistry*, 32:389–394, 1993.

[55] W. K. Surewicz, T. M. Stepanik, A. G. Szabo, and H. H. Mantsch. *Lipid-induced changes in the secondary structure of snake venom cardiotoxins*. *J. Biol. Chem.*, 263:786–790, 1988.

[56] S. J. Prestrelski, D. M. Byler, and M. N. Liebman. *Comparison of various molecular forms of Bovine Trypsin; correlation of infrared spectra with x-ray crystal structures. Biochemistry*, 30:133–143, 1991.

[57] C. L. Wilder, A. D. Friedrich, R. O. Potts, G. O. Daumy, and M. L. Francoeur. *Secondary Structural Analysis of Two Recombinant Murine Proteins, Interleukins 1α and 1β: Is Infrared Spectroscopy Sufficient to Assign Structure? Biochemistry*, 31:27–31, 1992.

[58] J. Trewhella, W. K. Liddle, D. B. Heidorn, and N. Strynadka. *Calmodulin and troponin C structures studied by Fourier transform infrared spectroscopy: effects of Ca^{2+} and Mg^{2+} binding. Biochemistry*, 28:1294–1301, 1989.

[59] M. Jackson and H. H. Mantsch. *Protein secondary structure from FT-IR spectroscopy: correlation with dihedral angles from three-dimensional Ramachandran plots. Can. J. Chem.*, 69:1639–1642, 1991.

[60] D. F. Kennedy, M. Crisma, C. Toniolo, and D. Chapman. *Studies of peptides forming 3_{10} and α-helices and β-bend ribbon structures in organic solution and in model biomembranes by Fourier transform infrared spectroscopy. Biochemistry*, 30:6541–6548, 1991.

[61] S. J. Prestrelski, D. M. Byler, and M. P. Thompson. *Infrared spectroscopic discrimination between α- and 3_{10}-helices in globular proteins. Re-examination of Amide I infrared bands of α-lactalbumin and their assignment to secondary structures. Int. J. Peptide Protein Res.*, 37:508–512, 1991.

[62] S. M. Miick, G. V. Martinez, W. R. Fiori, A. P. Todd, and G. L. Millhauser. *Short alanine-based peptides may form 3_{10}-helices and not α-helices in aqueous solution. Nature*, 359:653–655, 1992.

[63] D. M. Byler and H. Susi. *Examination of the secondary structure of proteins by deconvolved FTIR spectra. Biopolymers*, 25:469–487, 1986.

[64] Y. N. Chirgadze and N. A. Nevskaya. *Infrared spectra and resonance interaction of Amide I vibration of the parallel-chain pleated sheet. Biopolymers*, 15:627–636, 1976.

[65] H. Susi and D. M. Byler. *Fourier transform infrared study of proteins with parallel β-chains. Archives of Biochemistry and Biophysics*, 258:465–469, 1987.

[66] R. Khurana and A. L. Fink. *Do parallel β-helix proteins have a unique Fourier transform infrared spectrum? Biophysics*, 78:994–1000, 2000.

[67] H. Torii and M. Tasumi. *Model calculations on the Amide I infrared bands of globular proteins. J. Chem. Phys.*, 96:3379–3387, 1992.

[68] J. L. R. Arrando, F. J. Blanco, L. Serrano, and F. M. Goñi. *Infrared evidence of a β-hairpin peptide structure in solution. FEBS Letters*, 384:35–37, 1996.

[69] J. R. Lakowicz. *Principles of Fluorescence Spectroscopy*. Springer, 2006.

[70] M. R. Eftink. *Intrinsic Fluorescence of Proteins*. In J. R. Lakowicz, editor, *Topics in Fluorescence Spectroscopy, Volume 6: Protein Fluorescence*. Springer, 2000.

[71] E. A. Burstein, N. S. Vedenkina, and M. N. Ivkova. *Fluorescence and the location of tryptophan residues in protein molecules. Photochemistry and photobiology*, 18:263–279, 1973.

[72] D. V. Waterhous and W. C. Johnson. *Importance of Environment in Determining Secondary Structure in Proteins. Biochemistry*, 33:2121–2128, 1994.

[73] R. Rajan and P. Balaram. *A model for the interaction of trifluoroethanol with peptides and proteins. Int. J. Peptide Protein Res.*, 48:328–336, 1996.

[74] L. Zhong, R. J. Putnam, W. C. Johnson, and A. G. Rao. *Design and synthesis of amphipathic antimicrobial peptides*. *Int. J. Peptide Protein Res.*, 45:337–347, 1995.

[75] M. Eilers, S. Hwang, and G. Schatz. *Unfolding and refolding of a purified precursor protein during import into isolated mitochondria*. *EMBO J.*, 7:1139–1145, 1988.

[76] H. H. J. de Jongh, J. A. Killian, and B. de Kruijff. *A water-lipid interface induces a highly dynamic folded state in apocytochrome c and cytochrome c, which may represent a common folding intermediate*. *Biochemistry*, 31:1636–1643, 1992.

[77] T. Endo and G. Schatz. *Latent membrane perturbation activity of a mitochondrial precursor protein is exposed by unfolding*. *EMBO J.*, 7:1153–1158, 1988.

[78] V. E. Bychkova, A. E. Dujsekina, S. I. Klenin, E. I. Tiktopulo, V. N. Uversky, and O. B. Ptitsyn. *Molten Globule-Like State of Cytochrome c under Conditions Simulation Those Near the Membrane Surface*. *Biochemistry*, 35:6058–6063, 1996.

[79] V. E. Bychkova and O. B. Ptitsyn. *The molten globule in vitro and in vivo*. *Chemtracts: Biochem. Mol. Biol.*, 4:133–163, 1993.

[80] O. B. Ptitsyn, V. E. Bychkova, and V. N. Uversky. *Kinetic and Equilibrium Folding Intermediates*. *Philos. Trans. R. Soc. London, B*, 348:35–41, 1995.

[81] L. A. Munishkina, C. Phelan, V. N. Uversky, and A. L. Fink. *Conformational behaviour and aggregation of alpha-synuclein in organic solvents: modeling the effects of membranes. Biochemistry*, 42:2720–2730, 2003.

[82] D. C. Montefiori. *Retroviral Immunology: Immune Response and Restoration*, chapter HIV-specific neutralizing antibodies. Humana Press, Totowa, NJ., 2001.

[83] A. Trkola, A. P. Pomales, H. Yuan, B. Korber, P. J. Maddon, and G. Allaway. *Cross-clade neutralization of primary isolates of HIV-1 by human monoclonal antibodies and tetrameric CD4-IgG*. *J. Virol.*, 69:6609–6617, 1995.

[84] T. Muster, F. Steindl, M. Purtscher, A. Trkola, A. Klima, and G. Himmler. *A conserved neutralizing epitope on gp41 of human immunodeficiency virus type 1*. *J. Virol.*, 67:6642–6647, 1993.

[85] M. B. Zwick, A. K. Labrjin, M. Wang, C. Spenlehauer, E. O. Saphire, J. M. Binley, J. P. Moore, and G. Stiegler. *Broadly neutralizing antibodies targeted to the membrane-proximal external region of human immunodeficiency virus type 1 glycoprotein gp41*. *J. Virol.*, 75:10892–10905, 2001.

[86] M. Montero, N. E. van Houten, X. Wang, and J. K. Scott. *The membrane-proximal external region of the human immunodeficiency virus type 1 envelope: dominant site of antibody neutralization and target for vaccine design. Microbiol Mol Biol Rev*, 72:54–84, 2008.

[87] G. Ofek, M. Tang, A. Sambor, H. Katinger, J. R. Mascola, R. Wyatt, and P. D. Kwong. *Structure and Mechanistic Analysis of the Anti-Human Immunodeficiency Virus Type 1 Antibody 2F5 in Complex with Its gp41 Epitope*. *J. Virol.*, 78:10724–10737, 2004.

[88] K. Salzwedel, J. T. West, and E. Hunter. *A conserved tryptophan-rich motif in the membrane-proximal region of the human immunodeficiency virus type 1 gp41 ectodomain is important for env-mediated fusion and virus infectivity*. *J. Virol.*, 73:2469–2480, 1999.

[89] X. Liang, S. Munshi, J. Shendure, G. Mark, M. E. Davies, and D. C. Freed. *Epitope insertion into variable loops of HIV-1 gp120 as a potential means to improve immunogenicity of viral envelope proteins*. *Vaccine*, 17:2862–2872, 1999.

[90] Y. Lu, Y. Xiao, J. Ding, M. P. Dierich, and Y.-H. Chen. *Multiepitope vaccines intensively increased levels of antibodies recognizing three neutralizing epitopes on human immunodeficiency virus-1 envelope protein*. *Scand. J. Immunol.*, 51:497–501, 2000.

[91] Z. Biron, S. Khare, A. O. Samson, Y. Hayek, F. Naider, and J. Anglister. *A Monomeric 3_{10}-Helix Is Formed in Water by a 13-Residue Peptide Representing the Neutralizing Determinant of HIV-1 on gp41*. *Biochemistry*, 41:12687–12696, 2002.

[92] Z. Ahmed and S. Asher. *UV Resonance Raman Investigation of a 3_{10}-Helical Peptide Reveals a Rough Energy Landscape. Biochemistry*, 45:9068–9073, 2006.

[93] P. R. Tulip, C. R. Gregor, R. Z. Troitzsch, G. J. Martyna, E. Cerasoli, G. Tranter, and J. Crain. *Conformational plasticity in an HIV-1 antibody epitope*. *J. Phys. Chem. B.*, 114:7942–7950, 2010.

[94] J. Coutant, H. Yu, M. J. Clement, A. Alfsen, F. Toma, P. A. Curmi, and M. Bomsel. *Both lipid environment and pH are critical for determining physiological solution structure of 3-D-conserved epitopes of the HIV-1 gp41-MPER peptide P1. Faseb J*, 22:4338–4351, 2008.

[95] Z. Y. Sun, K. J. Oh, M. Kim, J. Yu, V. Brusic, L. Song, Z. Qiao, J. H. Wang, G. Wagner, and E. L. Reinherz. *HIV-1 broadly neutralizing antibody extracts its epitope from a kinked gp41 ectodomain region on the viral membrane*. *Immunity*, 28:52–63, 2008.

[96] D. J. Schibli, R. C. Montelaro, and H. J. Vogel. *The membrane-proximal tryptophan-rich region of HIV glycoprotein, gp41, forms a well-defined helix in dodecylphosphocholine micelles*. *Biochemistry*, 40:9570–9578, 2001.

[97] R. Pejchal, J. S. Gach, F. M. Brunel, R. M. Cardoso, R. L. Stanfield, P. E. Dawson, D. R. Burton, M. B. Zwick, and I. A. Wilson. *A conformational switch in human immunodeficiency virus gp41 revealed by the structures of overlapping epitopes recognized by neutralizing antibodies. J. Virol.*, 83:8451–8462, 2009.

[98] T. Imamura and K. Konishi. *Interaction of Tryptophan Dipeptides with Sodium Dodecyl Sulfate Micelles*. *J. Colloid Interface Sci.*, 198:300–307, 1998.

[99] M. E. Tuckerman, D. A. Yarne, S. O. Samuelson, A. L. Hughes, and G. J. Martyna. *Exploiting multiple levels of parallelism in Molecular Dynamics based calculations via modern techniques and software paradigms on distributed memory computers*. *Comput. Phys. Commun.*, 128:333–376, 2000.

[100] S. O. Samuelson, D. J. Tobias, and G. J. Martyna. *Modern Computational Methodology Applied to the Simulation of Blocked Trialanine Peptide in Vavuo, Water Clusters and Bulk Water*. *J. Phys. Chem. B*, 101:7592–7603, 1997.

[101] A. M. T. Martins Do Canto, A. J. Palace Carvalho, J. P. Prates Ramalho, and L. M. S. Loura. *T-20 and T-1249 HIV fusion inhibitors' structure and conformation in solution: a molecular dynamics study*. *Journal of Peptide Science*, 14:442–447, 2008.

[102] M. Karplus and J. A. McCammon. *Dynamics of proteins: Elements and function*. *Ann. Rev. Biochem*, 53:263–300, 1983.

[103] H. Frauenfelder, S. G. Sligar, and P. G. Wolynes. *The energy landscapes and motions of proteins*. *Science*, 254:1598–1603, 1991.

[104] J. N. Onuchic, Z. Luthey-Schulten, and P. G. Wolynes. *Theory of protein folding: The energy landscape perpsective*. *Ann. Rev. Phys. Chem.*, 48:545–600, 1997.

[105] M. Lapelosa, E. Gallicchio, G. F. Arnold, E. Arnold, and R. M. Levy. *In Silico vaccine design based on molecular simulations of rhinovirus chimeras presenting HIV-1 gp41 epitopes*. *J. Mol. Biol.*, 385:675–691, 2009.

[106] J. R. Forman, Z. T. Yew, S. Qamar, R. N. Sandford an E. Paci, and J. Clarke. *Non-Native Interactions Are Critical for Mechanical Strength in PKD Domains*. *Structure*, 17:1582–1590, 2009.

[107] M. Heinig and D. Frishman. *STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins*. *Nucleic Acids Research*, 32:W500–W502, 2004.

[108] J. W. Nelson and N. R. Kallenbach. *Stabilization of the ribonuclease S-peptide α-helix by trifluoroethanol*. *Poteins Struct. Funct. Genet.*, 1:211–217, 1986.

[109] J. W. Nelson and N. R. Kallenbach. *Persistence of the α-helix stop signal in the S-peptide in trifluoroethanol solutions*. *Biochemistry*, 28:5256–5261, 1989.

[110] G. Merutka and E. Stellwagen. *Analysis of peptides for helical prediction*. *Biochemistry*, 28:352–357, 1989.

[111] J. S. Balcerski, E. S. Pysh, G. M. Bonora, and C. Toniolo. *Vacuum ultraviolet circular dichroism of beta-forming alkyl oligopeptides*. *J. Am. Chem. Soc.*, 98:3470–3473, 1976.

[112] M. M. Kelly, E. S. Pysh, G. M. Bonora, and C. Toniolo. *Vacuum ultraviolet circular dichroism of protected homooligomers derived from L-leucine*. *J. Am. Chem. Soc.*, 99:3264–3266, 1977.

[113] U. Narayanan, T. A. Keiderling, G. M. Bonora, and C. Toniolo. *Vibrational circular dichroism of polypeptides. VII: Film and solution studies of β-sheet-forming homooligopeptides*. *J. Am. Chem. Soc.*, 108:2431–2437, 1986.

[114] L. Zhong and Jr. W. Curtis Johnson. *Environment affects amino acid preference for secondary structure*. *Proc. Nat'l. Acad. Sci. USA*, 89:4462–4465, 1992.

[115] C. R. D. Lancaster, P. K. Mishra, D. W. Hughes, S. A. St.-Pierre, A. A. Bothner-By, and R. M. Espand. *Mimicking the Membrane-Mediated Conformation of Dynorphin A-(1-13)-peptide: Circular Dichroism and Nuclear Magnetic Resonance Studies in Methanolic Solution*. *Biochemistry*, 30:4715–4726, 1991.

[116] J. D. Hirst and C. L. Brooks III. *Helicity, circular dichroism and molecular dynamics of proteins*. *J. Mol. Biol.*, 243:173–178, 1994.

[117] E. K. O'Shea, R. Rutkowski, and P. S. Kim. *Evidence that the leucine zipper is a coild coil*. *Science*, 243:538–542, 1989.

[118] K. D. Wilkinson and A. N. Mayer. *Alcohol-induced conformational changes of ubiquitin*. *Arch. Biochem. Biophys.*, 250:390–399, 1986.

[119] E. Dufour, C. Bertrand-Harb, and T. Haertlé. *Reversible effects of medium dielectric constant on structural transformation of β-lactoglobulin and its retinol binding*. *Biopolymers*, 33:589–598, 1993.

[120] Jr. L. Otvos, G. I. Szendrei, V. M.-Y. Lee, and H. H. Mantsch. *Human and rodent Alzheimer β-amyloid peptides acquire distinct conformations in membrane-mimicking solvents*. *Eur. J. Biochem*, 211:249–257, 1993.

[121] M. Buck. *Trifluoroethanol and colleagues: cosolvents come of age. Recent studies with peptides and proteins. Q. Rev. Biophys*, 31:297–355, 1998.

[122] C.-S. C. Wu and J. T. Yang. *Sequence-dependent conformations of short polypeptides in a hydrophobic environment*. *Mol. Cell. Biochem.*, 40:109–122, 1981.

[123] K. V. Pervushin, V. Yu. Orekhov, A. I. Popov, L. Yu. Musina, and A. S. Arseniev. *3D structure of (1-71)bacterioopsin solubilized in organic mixture and SDS micelles determined by ^{1}H- ^{15}N NMR*. *Eur. J. Biochem.*, 219:571–583, 1994.

[124] D. K. Chang, S. F. Cheng, and W. J. Chien. *The amino-terminal fusion domain peptide of human immunodeficiency virus type 1 gp41 inserts into the sodium dodecyl sulfate micelle primarily as a helix with a conserved glycine at the micelle-water interface*. *J. Virol.*, 71:6593–6602, 1997.

[125] C. H. Papavoine, R. N. Konings, C. W. Hilbers, and F. J. Van de Ven. *Location of M13 coat protein in sodium dodecyl sulfate micelles as determined by NMR. Biochemistry*, 33:12990–12997, 1994.

[126] V. Chupin, J. A. Killian, J. Breg, H. H. de Jongh, R. Boelens, R. Kaptein, and B. de Kruijff. *PhoE signal peptide inserts into micelles as a dynamic helix-break-helix structure, which is modulated by the environment. A two-dimensional ^1H NMR study. Biochemistry*, 34:11617–11624, 1995.

[127] H. W. van de Hooven, C. C. Doeland, M. Van De Kamp, R. N. Konings, C. W. Hilbers, and F. J. Van De Ven. *Three-Dimensional Structure of the Lantibiotic Nisin in the Presence of Membrane-Mimetic Micelles of Dodecylphosphocholine and of Sodium Dodecylsulphate. Eur. J. Biochem.*, 235:382–393, 1996.

[128] N. Sreerama and R. W. Woody. *Computation and analysis of protein circular dichroism spectra. Numerical Computer Methods, Part D*, 383:318–351, 2004.

[129] L. W. Tinoco, Jr. A. da Silva, A. Leite, A. P. Valente, and F. C. L. Almeida. *NMR Structure of PW2 Bound to SDS Micelles. J. Biol. Chem.*, 277:36351–36356, 2002.

[130] G. A. Dykes, R. E. W. Hancock, and J. W. Hastings. *Structural Variations in Nisin Associated with Different Membrane Mimicking and pH Environments. Biochem. Biophys. Res. Commun.*, 247:723–727, 1998.

[131] X. Shen, S. M. Dennison, P. Liu, F. Gao, F. Jaeger, D. C. Montefiori, L. Verkoczy, B. F. Haynes, S. M. Alam, and G. D. Tomaras. *Prolonged exposure of the HIV-1 gp41 membrane proximal region with L669S substitution. Proc. Nat. Acad. Sci. USA*, 107:5972–5977, 2010.

[132] E. Noah, Z. Biron, F. Naider, B. Arshava, and J. Anglister. *The Membrane Proximal External Region of the HIV-1 Envelope Glycoprotein gp41 Contributes to the Stabilization of the Six-Helix Bundle Formed with a Matching N' Peptide. Biochemistry*, 47:6782–6792, 2008.

[133] A. K. Bellamy-McIntyre, C. S. Lay, S. Baar, A. L. Maerz, G. H. Talbo, H. E. Drummer, and P. Poumbourios. *Functional Links between the Fusion Peptide-proximal Polar Segment and Membrane-proximal Region of Human Immunodeficiency Virus gp41 in Distinct Phases of Membrane Fusion. J. Biol. Chem.*, 282:23104–23116, 2007.

[134] F. Naider and J. Anglister. *Peptides in the treatment of AIDS. Curr. Opin. Struct. Biol.*, 19:473–482, 2009.

[135] M. J. Root, M. S. Kay, and P. S. Kim. *Protein Design of an HIV-1 Entry Inhibitor. Science*, 291:884–888, 2001.

[136] K. Champagne, A. Shishido, and M. J. Root. *Interactions of HIV-1 Inhibitory Peptide T20 with the gp41 N-HR Coiled Coil. J. Biol. Chem.*, 284:3619–3627, 2009.

[137] V. D. Trivedi, S. F. Cheng, C. W. Wu, R. Karthikeyan, C. J. Chen, and D. K. Chang. *The LLSGIV stretch of the N-terminal region of HIV-1 gp41 is critical for binding to a model peptide, T20. Protein Eng.*, 16:311–317, 2003.

[138] Z. Biron, S. Khare, S. R. Quadt, Y. Hayek, F. Naider, and J. Anglister. *The 2F5 Epitope Is Helical in the HIV-1 Entry Inhibitor T-20. Biochemistry*, 44:13602–13611, 2005.

[139] Y. Kliger, S. A. Gallo, S. G. Peisajovich, I. Munoz-Barroso, S. Avkin, R. Blumenthal, and Y. Shai. *Mode of Action of an Antiviral Peptide from HIV-1. J. Biol. Chem.*, 276:1391–1397, 2001.

[140] C. M. Finnegan, W. Berg, G. K. Lewis, and A. L. DeVico. *Antigenic Properties of the Human Immunodeficiency Virus Transmembrane Glycoprotein during Cell-Cell Fusion. J. Virol.*, 76:12123–12134, 2002.

[141] F. M. Brunel, M. B. Zwick, R. M. Cardoso, J. D. Nelson, I. A. Wilson, D. R. Burton, and P. E. Dawson. *Structure-Function Analysis of the Epitope for 4E10, a Broadly Neutralizing Human Immunodeficiency Virus Type 1 Antibody*. *J. Virol.*, 80:1680–1687, 2006.

[142] L. M. Gordon, P. W. Mobley, W. Lee, S. Eskandari, Y. N. Kaznessis, M. A. Sherman, and A. J. Waring. *Conformational mapping of the N-terminal peptide of HIV-1 gp41 in lipid detergent and aqueous environments using ^{13}C-enhanced Fourier transform infrared spectroscopy*. *Protein Sci.*, 13:1012–1030, 2004.

[143] F. B. Pereira, F. M. Goni, and J. L. Nieva. *Membrane Fusion Induced by the HIV Type 1 Fusion Peptide: Modulation by Factors Affecting Glycoprotein 41 Activity and Potential Anti-HIV Compounds*. *AIDS Res. Hum. Retrovir.*, 13:1203–1211, 1997.

[144] S. G. Peisajovich and Y. Shai. *HIV gp41: A Viral Membrane Fusion Machine*. In W. B. Fischer, editor, *Viral Membrane Proteins: Structure, Function, and Drug Design*. Springer, 2005.

[145] C. S. Lay, K. A. Wilson, B. Kobe, B. E. Kemp, H. E. Drummer, and P. Poumbourios. *Expression and biochemical analysis of the entire HIV-2 gp41 ectodomain: determinants of stability map to N- and C-terminal sequences outside the 6-helix bundle core*. *FEBS Lett.*, 567:183–188, 2004.

[146] W. Shu, H. Ji, and M. Lu. *Interactions between HIV-1 gp41 Core and Detergents and Their Implications for Membrane Fusion*. *J. Biol. Chem.*, 275:1839–1845, 2000.

[147] S. G. Peisajovich, L. Blank, R. F. Epand, R. M. Epand, and Y. Shai. *On the Interaction Between gp41 and Membranes: The Immunodominant Loop Stabilizes gp41 Helical Hairpin Conformation*. *J. Mol. Biol.*, 326:1489–1501, 2003.

[148] R. M. Cardoso, M. B. Zwick, R. L. Stanfield, R. Kunert, J. M. Binley, H. Katinger, D. R. Burton, and I. A. Wilson. *Broadly Neutralizing Anti-HIV Antibody 4E10 Recognizes a Helical Conformation of a Highly Conserved Fusion-Associated Motif in gp41*. *Immunity*, 22:163–173, 2005.

[149] N. Huarte, M. Lorizate, R. Maeso, R. Kunert, R. Arranz, J. M. Valpuesta, and J. L. Nieva. *The Broadly Neutralizing Anti-Human Immunodeficiency Virus Type 1 4E10 Monoclonal Antibody Is Better Adapted to Membrane-Bound Epitope Recognition and Blocking than 2F5*. *J. Virol.*, 82:8986–8996, 2008.

[150] G. R. Matyas, Z. Beck, N. Karasavvas, and C. R. Alving. *Lipid binding properties of 4E10, 2F5, and WR304 monoclonal antibodies that neutralize HIV-1*. *Biochim. Biophys. Acta.*, 1788:660–665, 2009.

[151] G. Ofek, K. McKee, Y. Yang, Z. Y. Yang, J. Skinner, F. J. Guenaga, R. Wyatt, M. B. Zwick, G. J. Nabel, J. R. Mascola, and P. D. Kwong. *Relationship between Antibody 2F5 Neutralization of HIV-1 and Hydrophobicity of Its Heavy Chain Third Complementarity-Determining Region*. *J. Virol.*, 84:2955–2962, 2010.

[152] A. S. Veiga, L. K. Pattenden, J. M. Fletcher, M. A. Castanho, and M. I. Aguilar. *Interactions of HIV-1 Antibodies 2F5 and 4E10 with a gp41 Epitope Prebound to Host and Viral Membrane Model Systems*. *ChemBioChem*, 10:1032–1044, 2009.

[153] M. Vila-Perello, A. Sanchez-Vallet, F. Garcia-Olmedo, A. Molina, and D. Andreu. *Structural Dissection of a Highly Knotted Peptide Reveals Minimal Motif with Antimicrobial Activity*. *J. Biol. Chem.*, 280:1661–1668, 2005.

[154] E. Glukhov, M. Stark, L. L. Burrows, and C. M. Deber. *Basis for Selectivity of Cationic Antimicrobial Peptides for Bacterial Versus Mammalian Membranes*. *J. Biol. Chem.*, 280:33960–33967, 2005.

[155] B. J. H. Kuipers and H. Gruppen. *Prediction of Molar Extinction Coefficients of Proteins and Peptides Using UV Absorption of the Constituent Amino Acids at 214 nm To Enable Quantitative Reverse Phase High-Performance Liquid Chromatography-Mass Spectrometry Analysis*. *J. Agric. Food Chem.*, 55:5445–5451, 2007.

[156] C. R. Gregor, E. Cerasoli, P. R. Tulip, M. G. Ryadnov, G. J. Martyna, and J. Crain. *Autonomous folding in the membrane proximal HIV peptide gp41$_{659-671}$: pH tuneability at micelle interfaces*. *Phys. Chem. Chem. Phys.*, 13:127–135, 2011.

[157] W. Humphrey, A. Dalke, and K. Schulten. *VMD - Visual Molecular Dynamics*. *J. Molec. Graphics*, 14.1:33–38, 1996.

[158] G. B. Melikyan, R. M. Markosyan, H. Hemmati, M. K. Delmedico, D. M. Lambert, and F. S. Cohen. *Evidence That the Transition of HIV-1 Gp41 into a Six-Helix Bundle, Not the Bundle Configuration, Induces Membrane Fusion*. *J. Cell Biol.*, 151:413–423, 2000.

[159] S. M. Alam, M. Morelli, S. M. Dennison, H. X. Liao, R. Zhang, S. M. Xia, S. Rits-Volloch, L. Sun, S. C. Harrison, B. F. Haynes, and B. Chen. *Role of HIV membrane in neutralization by two broadly neutralizing antibodies*. *Proc. Nat. Acad. Sci. USA*, 106:20234–20239, 2009.

[160] E. Cerasoli, J. Ravi, C. R. Gregor, R. Hussain, G. Siligardi, G. J. Maryna, J. Crain, and M. G. Ryadnov. *Primitive Membrane Mediated Regulation in Free Peptides of HIV-1 gp41: Implications for the Hemifusion Phase*. Submitted.

[161] T. Pott, J. Dufourcq, and E. J. Dufourc. *Fluid or gel phase lipid bilayers to study peptide-membrane interactions? Eur. Biophys. J.*, 25:55–59, 1996.

[162] I. Munoz-Barroso, K. Salzwedel, E. Hunter, and R. Blumenthal. *Role of the membrane-proximal domain in the initial stages of human immunodeficiency virus type 1 envelope glycoprotein-mediated membrane fusion*. *J. Virol.*, 73:6089–6092, 1999.

[163] M. G. Ryadnov, G. V. Mukamolova, A. S. Hawrani, J. Spencer, and R. Platt. *RE?Coil: An Antimicrobial Peptide Regulator*. *Angew. Chem. Int. Ed.*, 48:9676–9679, 2009.

[164] A. Bitler, N. Lev, Y. Fridmann-Sirkis, L. Blank, S. R. Cohen, and Y. Shai (2010). *Kinetics of interaction of HIV fusion protein (gp41) with lipid membranes studied by real-time AFM imaging*. *Ultramicroscopy*, 110:694–700, 2010.

[165] N. Lev, Y. Fridmann-Sirkis, L. Blank, A. Bitler, R. F. Epand, R. M. Epand, and Y. Shai. *Conformational Stability and Membrane Interaction of the Full-Length Ectodomain of HIV-1 gp41: Implication for Mode of Action*. *Biochemistry*, 48:3166–3175, 2009.

[166] L. A. Cole. *Hyperglycosylated hCG*. *Placenta*, 28:977–986, 2007.

[167] J. F. O'Connor, S. Birken, J. W. Lustbader, A. Krichevsky, Y. Chen, and R. E. Canfield. *Recent Advances in the Chemistry and Immunochemistry of Human Chorionic Gonadotropin: Impact on Clinical Measurements*. *Endocr. Rev.*, 15:650–683, 1994.

[168] C. M. Sturgeon and E. J. McAllister. *Analysis of hCG: clinical applications and assay requirements*. *Ann. Clin. Biochem.*, 35:460–491, 1998.

[169] A. M. Gronowski and D. G. Grenache. *Characterization of the hCG Variants Recognized by Different hCG Immunoassays: An Important Step Toward Standardization of hCG Measurements*. *Clin. Chem.*, 55:1447–1449, 2009.

[170] U. H. Stenman. *Immunoassay Standardization: Is It Possible, Who Is Responsible, Who Is Capable? Clin. Chem.*, 47:815–820, 2001.

[171] M. Ozturk, D. Bellet, L. Manil, G. Hennen, R. Frydman, and J. R. Wands. *Physiological Studies of Human Chorionic Gonadotropin (hCG), αhCG, and βhCG as Measured by Specific Monoclonal Immunoradiometric Assays*. *Endocrinology*, 120:549–558, 1987.

[172] M. Ozturk, R. Berkowitz, D. Goldstein, D. Bellet, and J. R. Wands. *Differential production of human chorionic gonadotropin and free subunits in gestational trophoblastic disease*. *Am. J. Obstet. Gynecol.*, 158:193–198, 1988.

[173] J. G. Pierce. *Eli Lilly Lecture: The Subunits of Pituitary Thyrotropin-Their Relationship to Other Glycoprotein Hormones*. *Endocrinology*, 89:1331–1344, 1971.

[174] J. G. Pierce and T. F. Parsons. *Glycoprotein Hormones: Structure and Function*. *Annu. Rev. Biochem.*, 50:465–495, 1981.

[175] T. Lund and P. J. Delves. *Immunological analysis of epitopes on hCG*. *Reviews of Reproduction*, 3:71–76, 1998.

[176] H. Wu, J. W. Lustbader, Y. Liu, R. E. Canfield, and W. A. Hendrickson. *Structure of human chorionic gonadotropin at 2.6 Å resolution from MAD analysis of the selenomethionyl protein*. *Structure*, 2:545–558, 1994.

[177] P. Berger, C. Sturgeon, J. M. Bidart, E. Paus, R. Gerth, M. Niang, A. Bristow, S. Birken, and U. H. Stenman. *The ISOBM TD-7 Workshop on hCG and Related Molecules*. *Tumor Biol.*, 23:1–38, 2002.

[178] M. M. Elliott, A. Kardana, J. W. Lustbader, and L. A. Cole. *Carbohydrate and peptide structure of the α- and β-subunits of human chorionic gonadotropin from normal and aberrant pregnancy and choriocarcinoma*. *Endocrine.*, 7:15–32, 1997.

[179] S. Birken, E. G. Armstrong, M. A. G. Kolks, G. M. Agosto L. A. Cole, A. Krichevsky, J. L. Vaitukaitis, and R. E. Canfield. *Structure of the human chorionic gonadotropin beta subunit fragment from pregnancy urine*. *Endocrinology*, 123:572–583, 1988.

[180] C. M. Sturgeon, P. Berger, J. M. Bidart, S. Birken, C. Burns, R. J. Norman, and U. H. Stenman. *Differences in Recognition of the 1st WHO International Reference Reagents for hCG-Related Isoforms by Diagnostic Immunoassays for Human Chorionic Gonadotropin*. *Clin. Chem.*, 55:1484–1491, 2009.

[181] H. Mitchell and M. J. Seckl. *Discrepancies between commercially available immunoassays in the detection of tumour-derived hCG*. *Molecular and cellular endocrinology*, 260-262:310–313, 2007.

[182] R. H. Meloen, W. C. Puijk, and W. M. M. Schaaper. *Immunology Methods Manual*, pages 982–988. Academic Press, 1997.

[183] J. Slootstra, W. Puijk, G. Ligtvoet, J. Langeveld, and R. Meloen. *Structural aspects of antibody-antigen interaction revealed through small random peptide libraries*. *Mol. Divers.*, 1:87–96, 1996.

[184] L. E. Valenti, M. B. Paci, C. P. De Pauli, and C. E. Giacomelli. *Infrared study of trifluoroacetic acid unpurified synthetic peptides in aqueous solution: Trifluoroacetic acid removal and band assignment*. *Anal. Biochem.*, 410:118–123, 2011.

[185] C. R. Gregor, E. Cerasoli, J. Schouten, J. Ravi, J. Slootstra, A. Horgan, G. J. Martyna, M. G. Ryadnov, P. Davis, and J. Crain. *Antibody recognition of a human chorionic gonadotropin epitope ($hCG\beta_{66-80}$) depends on local structure retained in the free peptide. J. Biol. Chem.*, 286:25016–25026, 2011.

[186] J. Bandekar and S. Krimm. *Vibrational analysis of peptides, polypeptides, and proteins: Characteristic amide bands of β-turns*. *Proc. Natl. Acad. Sci.*, 76:774–777, 1979.

[187] N. C. Seeman and A. M. Belcher. *Emulating biology; building nanostructures from the bottom up*. *Proc. Natl. Acad. Sci. USA*, 99:6451–6455, 2002.

[188] A. Aggeli, M. Bell, L. M. Carrick, C. W. G. Fishwick, R. Harding, P. J. Mawer, S. E. Radford, A. E. Strong, and N. Boden. *pH as a trigger of peptide β-sheet self-assembly and reversible switching between nematic and isotropic phases*. *J. Am. Chem. Soc.*, 125:9619–9628, 2003.

[189] T. C. Holmes, S. De Lacalle, X. Su, G. S. Liu, A. Rich, and S. G. Zhang. *Extensive neurite outgrowth and active synapse formation on self-assembling peptide scaffolds*. *Proc. Natl. Acad. Sci. USA*, 97:6728–6733, 2000.

[190] T. C. Holmes. *Novel peptide-based biomaterial scaffolds for tissue engineering*. *Trends Biotechnol.*, 20:16–21, 2002.

[191] N. A. Peppas, P. Bures, W. Leobandung, and H. Ichikawa. *Hydrogels in Pharmaceutical Formulations*. *Eur. J. Pharmaceutics and Biopharmaceutics*, 50:27–46, 2000.

[192] M. Sunde and C. C. F. Blake. *From the globular to the fibrous state: protein structure and structural conversion in amyloid formation*. *Q. Rev. Biophys.*, 31:1–39, 1998.

[193] A. Clark, S. B. Charge, M. K. Badman, D. A. MacArthur, and E. J. de Koning. *Islet amyloid polypeptide: actions and role in the pathogenesis of diabetes*. *Biochem. Soc. Trans.*, 24:594–599, 1996.

[194] J. W. Kelly. *The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways*. *Curr. Opin. Struct. Biol.*, 8:101–106, 1998.

[195] C. M. Dobson. *The structural basis of protein folding and its links with human disease*. *Philos. Trans. R. Soc. London B*, 356:133–145, 2001.

[196] J. I. Guijarro, M. Sunde, J. A. Jones, I. D. Campbell, and C. M. Dobson. *Amyloid fibril formation by an SH_3 domain*. *Proc. Natl. Acad. Sci. USA*, 95:4224–4228, 1998.

[197] F. Chiti, P. Webster, N. Taddei, A. Clark, M. Stefani, G. Ramponi, and C. M. Dobson. *Designing conditions for in vitro formation of amyloid protofilaments and fibrils*. *Proc. Natl. Acad. Sci. USA*, 96:3590–3594, 1999.

[198] M. Sunde and C. Blake. *The Structure of Amyloid Fibrils by Electron Microscopy and X-Ray Diffraction*. *Ad. Protein Chem.*, 50:123–159, 1997.

[199] S. H. Waterhouse and J. A. Gerrard. *Amyloid Fibrils in Bionanotechnology*. *Aust. J. Chem.*, 57:519–523, 2004.

[200] C. C. F. Blake, M. J. Geisow, S. J. Oatley, B. Rerat, and C. Rerat. *Structure of prealbumin: Secondary, tertiary and quaternary interactions determined by Fourier refinement at 1.8Å*. *J. Mol. Biol.*, 121:339–356, 1978.

[201] A. M. Damas and M. J. Saraiva. *TTR amyloidosisÐstructural features leading to protein aggregation and their implications on therapeutic strategies. J. Struct. Biol.*, 130:290–299, 2000.

[202] M. M. Sousa and M. J. Saraiva. *Neurodegeneration in familial amyloid polyneuropathy: from pathology to molecular signaling. Prog. Neurobiol.*, 71:385–400, 2003.

[203] A. Gustavsson, U. Engstrom, and P. Westermark. *Normal transthyretin and synthetic transthyretin fragments from amyloid-like fibrils in vitro*. *Biochem. Biophys. Res. Commun.*, 175:1159–1164, 1991.

[204] J. A. Jarvis, D. J. Craik, and M. C. J. Wilce. *X-Ray-Diffraction Studies of Fibrils Formed from Peptide Fragments of Transthyretin*. *Biochem. Biophys. Res. Commun.*, 192:991–998, 1993.

[205] C. P. Jaroniec, C. E. MacPhee, N. S. Astrof, C. M. Dobson, and R. G. Griffin. *Molecular conformation of a peptide fragment of transthyretin in an amyloid fibril*. *Proc. Natl. Acad. Sci. USA*, 99:16748–16753, 2002.

[206] C. P. Jaroniec, C. E. MacPhee, V. S. Bajaj, M. T. McMahon, C. M. Dobson, and R. G. Griffin. *High-resolution molecular structure of a peptide in an amyloid fibril determined by magic angle spinning NMR spectroscopy*. *Proc. Natl. Acad. Sci. USA*, 101:711–716, 2004.

[207] C. Dirix, F. Meersman, C. E. MacPhee, C. M. Dobson, and K. Heremans. *High Hydrostatic Pressure Dissociates Early Aggregates of TTR_{105115}, but not the Mature Amyloid Fibrils*. *J. Mol. Biol.*, 347:903–909, 2005.

[208] E. Paci, J. Gsponer, X. Salvatella, and M. Vendruscolo. *Molecular Dynamics Studies of the Process of Amyloid Aggregation of Peptide Fragments of Transthyretin*. *J. Mol. Biol.*, 340:555–569, 2004.

[209] H. L. Cole, J. M. D. Kalapothakis, G. Bennett, P. E. Barran, and C. E. MacPhee. *Characterizing Early Aggregates Formed by an Amyloidogenic Peptide by Mass Spectrometry*. *Angew. Chem. Int. Ed.*, 49:9448–9451, 2010.

[210] D.-W. Li, L. Han, and S. Huo. *Structural and Pathway Complexity of β-Strand Reorganization within Aggregates of Human Transthyretin(105-115) Peptide*. *J. Phys. Chem. B*, 111:5425–5433, 2007.

[211] M. F. Mesleh, J. M. hunter, A. A. Shvartsburg, G. C. Schatz, and M. F. Jarrold. *Structural Information from Ion Mobility Measurements: Effects of the Long-Range Potential*. *J. Phys. Chem.*, 100:16082–16086, 1996.

[212] J. A. Jarvis, A. Kirkpatrick, and D. J. Craik. *H1 NMR analysis of fibril-forming peptide fragments of transthyretin. Int. J. Pept. Protein Res.*, 44:388–398, 1994.

[213] S. A. Peterson, T. Klabunde, H. A. Lashuel, H. Purkey, J. C. Sacchettini, and J. W. Kelly. *Inhibiting transthyretin conformational changes that lead to amyloid fibril formation*. *Proc. Natl. Acad. Sci. USA*, 95:12956–12960, 1998.

[214] A.L. Jenkins, R.A. Larsen, and T.B. Williams. *Characterization of amino acids using Raman spectroscopy*. *Spectrochemica Acta Part A*, 61:1585–1594, 2005.

[215] K.J. Jalkanen, M. Elstner, and S. Suhai. *Amino acids and small peptides as building blocks for proteins: comparitive theoretical and spectroscopic studies*. *Journal of Molecular Structure*, 675:61–67, 2004.

[216] P. Bour and T.A. Keiderling. *Structure, spectra and the effects of twisting of β-sheet peptides. A density functional theory study*. *Journal of Molecular Structure*, 675:95–105, 2004.

[217] L. Pauling, R.B. Corey, and H.R. Branson. *The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain*. *Proc. Nat'l. Acad. Sci. USA*, 37:205–212, 1951.

[218] J. Donohue. *Hydrogen bonded helical conformations of the polypeptide chain*. *Proc. Nat'l Acad. Sci. USA*, 39:470–478, 1953.

[219] G.L. Millhauser. *Views of helical peptides: a proposal for the position of the 3_{10}-helix along the thermodynamic folding pathway*. *Biochemistry*, 34:3873–3877, 1995.

[220] W.T. Astbury. *Some problems in the X-ray analysis of the structure of animal hairs and other protein fibres*. *Trans. Faraday Soc.*, 29:193–205, 1933.

[221] J.S. Richardson. *β-sheet topology and the relatedness of proteins*. *Nature*, 268:495–500, 1977.

[222] C.L. Nesloney and J.W. Kelly. *Progress towards understanding β-sheet structure*. *Bioorganic & Medicinal Chemistry*, 4:739–766, 1996.

[223] D.N. Woolfson, P.A. Evans, E.G. Hutchinson, and J.M. Thornton. *Topological and stereochemical restrictions on β-sandwich protein structures*. *Protein Engineering*, 6:461–470, 1993.

[224] J. Kubelka and T.A. Keiderling. *Differentiation of β-sheet forming structures; ab-initio based simulations of IR absorrption and vibrational CD for model peptide and protein β-sheets*. *J. Am. Chem. Soc.*, 123:12048–12058, 2001.

[225] J.A. Smith and L.G. Pease. *Reverse turns in peptides and proteins*. *CRC Crit. Rev. Biochem*, 8:315–399, 1980.

[226] G.D. Rose, L.M. Gierasch, and J.A. Smith. *Turns in Peptides and Proteins*. In C. B. Anfinsen, J. T. Edsall, and F. M. Richards, editors, *Advances in Protein Chemistry, Vol. 37*. Academic Press Inc., 1985.

[227] T.S. Haque, J.C. Little, and S.H. Gellman. *Stereochemical requirements for β-hairpin formation: model studies with four-residue peptides and depsipeptides*. *J. Am. Chem. Soc.*, 118:6975–6985, 1996.

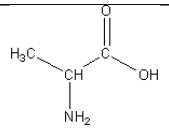
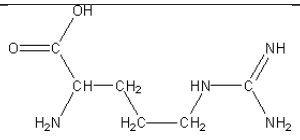
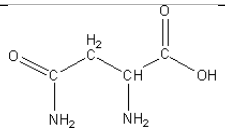
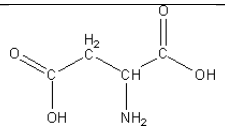
[228] P. Y. Chou and G. D. Fasman. *Conformational parameters for amino acids in helical, β-sheet, and random coil regions calculated from proteins*. *Biochemistry*, 13:211–222, 1974.

A

Peptides and Proteins

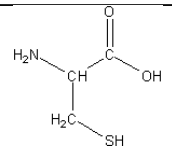
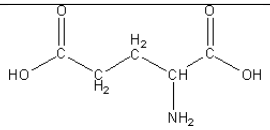
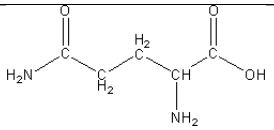
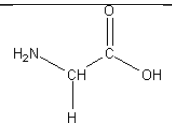
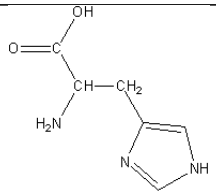
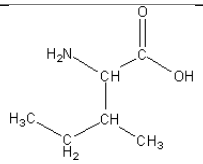
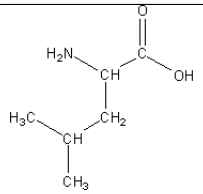
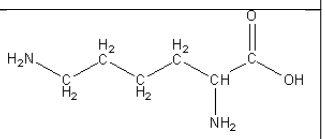
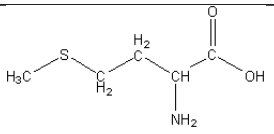
A.1 The Primary Structure of Peptides

All peptides and polypeptides are biopolymers made of amino acids, with 20 relevant α-amino acids (Table. A.1) available for protein synthesis [214]. The α-amino acids contain a carboxylic acid (-COOH) and an amino (-NH_2) functional group attached to the same carbon (known as the α-carbon). In addition to these groups, the amino acids can be distinguished by distinct R groups which are also attached to this carbon.

Name	3 (1)-letter	Polarity	Charge	R - pKa	Structure
Alanine	Ala (A)	non-polar	neutral		
Arginine	Arg (R)	polar	positive	12.48	
Asparagine	Asn (N)	polar	neutral		
Aspartic Acid	Asp (D)	polar	negative	3.90	

Continued on next page

Table A.1 – continued from previous page

Name	3 (1)-letter	Polarity	Charge	R - pKa	Structure
Cysteine	Cys (C)	non-polar	neutral	8.37	
Glutamic Acid	Glu (E)	polar	negative	4.07	
Glutamine	Gln (Q)	polar	neutral		
Glycine	Gly (G)	non-polar	neutral		
Histidine	His (H)	polar	neutral	6.04	
Isoleucine	Ile (I)	non-polar	neutral		
Leucine	Leu (L)	non-polar	neutral		
Lysine	Lys (K)	polar	positive	10.54	
Methionine	Met (M)	non-polar	neutral		

Continued on next page

Table A.1 – continued from previous page

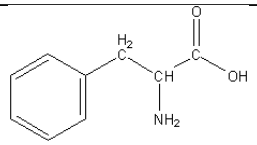
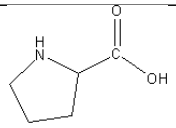
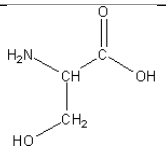
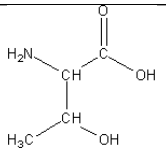
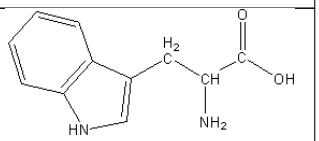
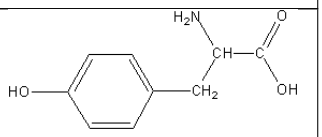
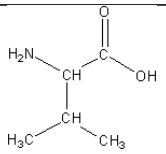
Name	3 (1)-letter	Polarity	Charge	R - pKa	Structure
Phenylalanine	Phe (F)	non-polar	neutral		
Proline	Pro (P)	non-polar	neutral		
Serine	Ser (S)	polar	neutral		
Threonine	Thr (T)	polar	neutral		
Tryptophan	Trp (W)	non-polar	neutral		
Tyrosine	Tyr (Y)	polar	neutral	10.46	
Valine	Val (V)	non-polar	neutral		

Table A.1: Amino Acids found in peptides, indicating their standard abbreviations, structure, side-chain polarity, side-chain charge at pH 7.4 and side-chain pKa [6]

These various peptides can be polymerised via a condensation reaction (Fig. A.1) between a carboxyl group of one acid with an amino group of another, to form a ‘peptide bond.’

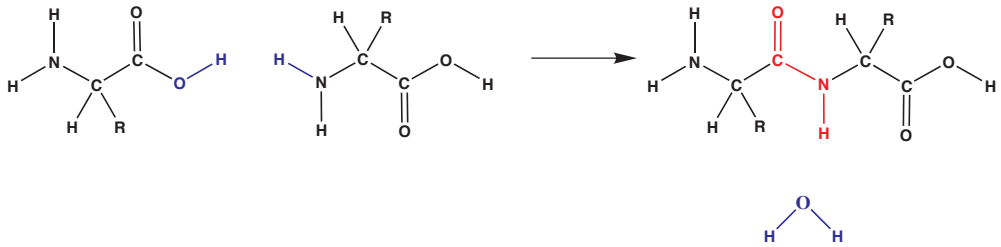


Figure A.1: Condensation reaction between two amino acids forming a peptide bond (red) and water (blue).

A.2 Secondary Structure

Due to the presence of the carbonyl group in the peptide bond, electron resonance stabilisation is permitted giving the peptide bond rigidity and a partial double-bond character. This results in the peptide bond being essentially planar. The shape of the peptide chain can be determined by three consecutive bond torsion angles (Fig. A.2), NH to C_α (phi, ϕ), C_α to C=O (psi, ψ) and C=O to NH (omega, ω). The latter relating to the rigid (planar) peptide bond and is usually fixed at 180°. This constraint limits the number of arrangements the peptide chain can take without distorting bonds or forcing atoms closer than their van der Walls radii will allow.

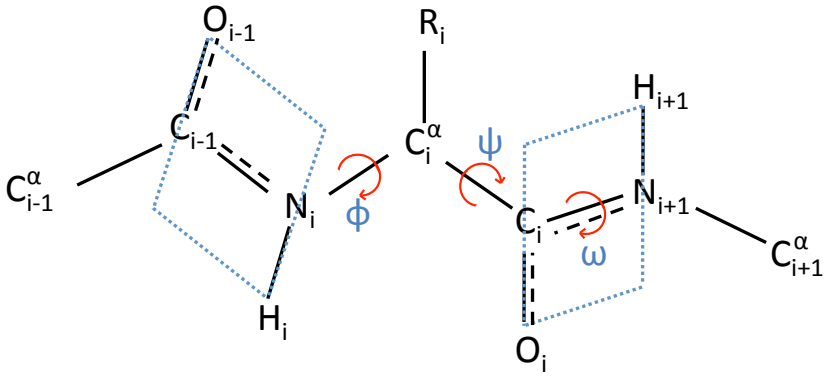


Figure A.2: The geometry of a peptide bond. The C(O)NH bond forms a single plane ($\omega = 180°$). The Ramachandran, or dihedral angles ϕ and ψ are variable: -180° < ϕ,ψ < 180°.

It is found that two general patterns emerge [215], which conform to the atomic geometry and dominate the structure of peptides. The first is the helical conformation (Fig. A.3a), which forms the most dominant uniform secondary structure element in proteins [216]. These helices are formed through intramolecular hydrogen bonding between the carbonyl oxygen of one pep-

tide bond and the hydrogen atom of another [47]. The α-helix was originally discovered by Pauling et. al. [217] and is regarded as being the most dominant of the helical conformations [9]. Other helical structures which had been proposed [218], were subsequently discovered and differentiated using their structural parameters (Tab. A.2). An important variation on the helical conformation is that of the 3_{10}-helix which is believed to be an intermediate in helix folding processes [216, 219]. It is a distinct class of helix and is usually found at the termini of regular α-helices.

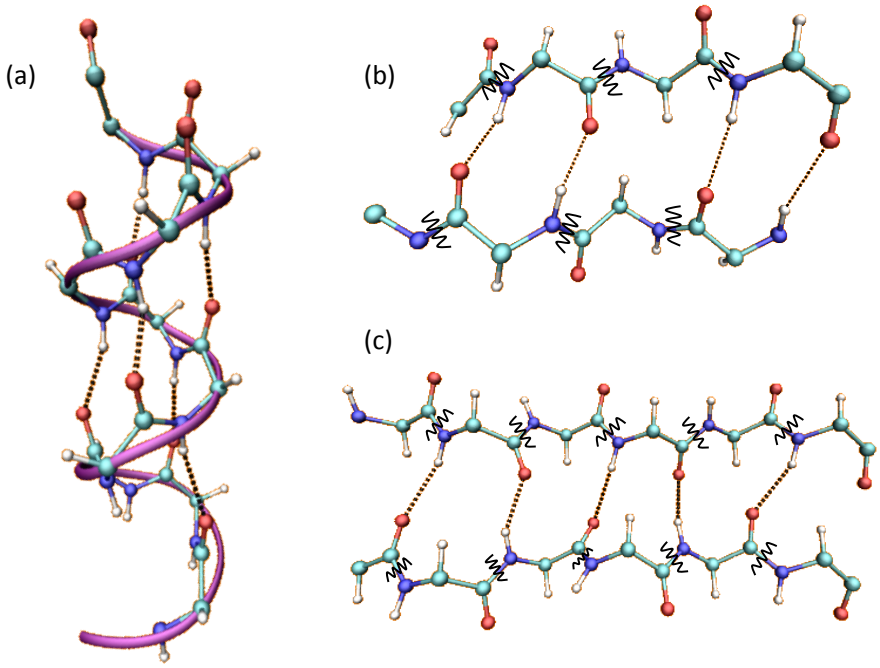


Figure A.3: Schematic showing hydrogen bonding in (a) α-helix, (b) Antiparallel β-sheet and (c) Parallel β-sheet (the latter two displaying boundaries of individual amino acids).

The other major conformation is that of an extended β-strand, originally derived from the conformation of fibrous proteins known as β-keratins [220]. Amino acid residues in this β-conformation have negative ϕ and positive ψ angles, typically -140° and +130° respectively [51], resulting in the peptide chain alternating in direction and a zigzag structure emerging. In a β-strand, it is not possible for intramolecular hydrogen bonding to occur; instead, strands running alongside one another pair up via hydrogen bonds to form extended pleated β-sheets. Depending on how these strands line up, β-sheets can be formed in either an antiparallel (Fig. A.3b) or a parallel manner (Fig. A.3c), with the strands running in an opposite or in the same direction to

Structure	n	m	ϕ	ψ	ω	H-bond
α_I-helix	3.60	13	-57°	-47°	180°	i to $i+4$
α_{II}-helix	3.60	13	-70°	-36°	180°	i to $i+4$
3_{10}-helix	2.99	10	-45°	-30°	176°	i to $i+3$
ω-helix	-4.00	13	64°	55°	175°	i to $i+4$
π-helix	4.40	16	-57°	-70°	180°	i to $i+5$
π-helix	-4.25	16	51°	74°	177°	i to $i+5$
2.2_7-helix	2.17	7	-78°	59°	180°	i to $i+2$

Table A.2: Parameters of various intramolecular helical structures [7]. **n**, number of residues per turn (positive for right-handed, negative for left-handed), **m**, number of atoms in hydrogen-bonded 'ring'.

one another. Parallel sheets are proposed to be less stable than their counterparts as they are usually buried within the hydrophobic interior of proteins [221], however the validity of this statement is not clear. Parallel sheets are also less common than antiparallel sheets due to their structural formation. β-turns (see later) serve to reverse the direction of a peptide chain allowing the formation of antiparallel sheets, parallel sheets however require two peptide segments which are not close in sequence to be brought together [222]. The segment connecting the strands may be a loop (above or below the plane of the sheet) or may exhibit secondary structure of its own (usually α-helix). Studies have shown that these long loops required to connect the strands are both kinetically and thermodynamically unfavourable [223]. In addition, finite β-sheets (as opposed to the ideal, infinite periodic models) are known to develop a twist; with the smaller the sheet, the larger the deviation [224].

As mentioned above, antiparallel sheets can be formed by reversing the direction of a polypeptide chain by 180°. This can be achieved through reverse turns which are characterised by their intramolecular hydrogen bonding which causes the reversal in direction [7]; an other definition of a turn is the close approach (< 7Å) of the two C^{α}. There are two major types of turns [225] which involve either four (β-turn) or three (γ-turn) successive amino acid residues; other turn types do exist. The conformation of β-turns can be described through the torsion angles of the second and third residues (Tab. A.3), while the γ-turns need only concern the torsion angles of the central residue [225, 226]. The majority of turns in folded proteins adopt type I and II β-turns [227], while type III β-turns are regarded as the building blocks of the 3_{10}-helix [7]. The simplest way to bring two antiparallel strands together is a short peptide segment between the

Designation	Residue 2 Phi	Psi	Residue 3 Phi	Psi	Comments
I	-60	-30	-90	0	Most common type
II	-60	120	80	0	
III	-60	-30	-60	-30	Like 3_{10}-helix
IV					
V	-80	80	80	-80	
VIa	-60	120	-90	0	2-3 peptide bond is cis, residue 3 is Pro.
VIb	-120	120	-60	0	2-3 peptide bond is cis, residue 3 is Pro.
VII					
VIII	-60	-30	-120	120	

Table A.3: Torsional angles of second and third residue in different β-turns. [8–10]

C-terminus of one strand and the N-terminus of the other. Referred to as a β-hairpin; the shortest common loop that this strand-loop-strand arrangement adopts is two residues, which constitute the middle two residues of a β-turn. While type I and type II β-turns are the dominant turn classes in folded proteins, they are rare in two-residue hairpins. Instead, β-hairpins are stabilised by the corresponding 'mirror image' type I' and type II'; these are diastereomers of type I and II β-turns where the torsional angles of the residues are of opposite sign, and as the chirality of the C^{α} is maintained, are not true mirror images. Furthermore, common type turns such as type I and type II are actually regarded as an impediment to the formation of the two-residue hairpin [227].

A.3 Structural Predictions

In order to predict the structure adopted by different peptides, a number of predictive programs have been designed which focus on the amino acid composition of the peptide [11, 228]. These classify individual amino acids as to whether they have the propensity to form (or break) an α-helix or β-sheet (Tab. A.4). Although these programs can be effective in predicting the structure to some extent, it has been shown experimentally that the periodicity of the polar and non-polar amino acids has a major impact on the secondary structure. Both α-helices and β-sheets can be amphiphilic and have well-defined hydrophobic (non-polar) and hydrophilic (polar) faces. These faces aid in the self-assembly of such structures by hydrophobic interactions, for example in water, the hydrophobic faces would be buried and the polar surfaces exposed. These two secondary structures have different periodicities and the pattern of hydrophobic (H) and polar

Helical Residues	$\mathbf{P}_\alpha$		β-Sheet Residues	$\mathbf{P}_\beta$	
E	1.53		M	1.67	
A	1.45	$\mathbf{H}_\alpha$	V	1.65	$\mathbf{H}_\beta$
L	1.34		I	1.6	
H	1.24		C	1.30	
M	1.20		Y	1.29	
Q	1.17	$\mathbf{h}_\alpha$	F	1.28	
W	1.14		Q	1.23	$\mathbf{h}_\beta$
V	1.14		L	1.22	
F	1.12		T	1.20	
K	1.07	$\mathbf{I}_\alpha$	W	1.19	
I	1.00		A	0.97	$\mathbf{I}_\beta$
D	0.98		R	0.90	
T	0.82		G	0.81	$\mathbf{i}_\beta$
S	0.79	$\mathbf{i}_\alpha$	D	0.80	
R	0.79		K	0.74	
C	0.77		S	0.72	
N	0.73	$\mathbf{b}_\alpha$	H	0.71	$\mathbf{b}_\beta$
Y	0.61		N	0.65	
P	0.59	$\mathbf{B}_\alpha$	P	0.62	
G	0.53		E	0.26	$\mathbf{B}_\beta$

Table A.4: $\mathbf{H}_\alpha$, strong α-helix former; $\mathbf{h}_\alpha$, α-helix former; $\mathbf{I}_\alpha$, weak α-former; $\mathbf{i}_\alpha$, α-indifferent; $\mathbf{b}_\alpha$, α-helix breaker; $\mathbf{B}_\alpha$, strong α-helix breaker. $\mathbf{H}_\beta$, strong β-sheet former; $\mathbf{h}_\beta$, β-sheet former; $\mathbf{I}_\beta$, weak β-sheet former; $\mathbf{i}_\beta$, β-sheet indifferent; $\mathbf{b}_\beta$, β-sheet breaker; $\mathbf{B}_\beta$, strong β-sheet breaker [11].

(P) amino acids would favour a particular secondary structure over the other; $HPHPHP$-type primary structures will favour β-strands, while $HPPHPPP$-type primary structures will favour α-helices.

B

Ewald Summation Method

For systems comprising of a large number of particles it is critical to avoid calculating the interaction between all pairs of particles due to the computational effort that this would require. This is usually resolved by truncation of interaction potentials at a cut-off r_c, where interactions between particles separated by a distance $>r_c$ are set to zero. If one was to truncate a potential at r_c, the contribution of the tail of the potential in 3D can be estimated (Eq. B.1)

$$\mathcal{U}_{tail} = \frac{N\rho}{2} \int_{r_c}^{\infty} dr\, U(r)\, 4\pi r^2 \tag{B.1}$$

A caveat for such a truncation is that this contribution will diverge unless the potential energy function associated with $U(r)$ decays faster than r^{-3}. In the case of long-range interactions such as the Coulombic interaction this truncation is therefore not possible (decays as r^{-1}) and one must therefore find an alternative method that will effectively handle long-ranged intermolecular interactions. One such method is that of the Ewald Summation method, originally devised by Ewald in 1921 in order to study the energetics of ionic crystals. This Appendix aims to present an outline of the Ewald sum for the calculation of long-range electrostatic interactions in a system with periodic boundary conditions.

B.1 Introduction

To begin with, one must consider a system of N charged particles (q_i) located within a cube of diameter L. One must assume that the sum of such charged particles results in the system being electrically neutral (i.e. $\sum_i q_i$ = 0) and the system is subjected to periodic boundary conditions. The coulombic contribution to the potential energy of such a system can be computed (Eq. B.2)

$$\mathcal{U}_{coulomb} = \frac{1}{2}\sum_{i=1}^{N} q_i \phi(r_i) \tag{B.2}$$

In order to compute this contribution, one makes use of the electrostatic potential at the position of the i^{th} charge, $\phi(r_i)$, (Eq. B.3). This summation is over all periodic images **n** and over all particles j. The prime indicates that the summation excludes the interaction of the particle with itself (i.e. j=i when **n**=0). Unfortunately this equation contains a poorly convergent sum and therefore cannot be used to calculate the electrostatic energy in a simulation.

$$\phi(r_i) = \sum_{j,\mathbf{n}}^{'} \frac{q_j}{\mid \mathbf{r}_{ij} + \mathbf{n}L \mid} \tag{B.3}$$

The trick associated with the Ewald sum is in converting the summation into two parts which both individually converge quickly (Fig. B.1). One can represent the charge density as a sum of δ-functions, with their contribution to the electrostatic potential decaying as 1/r. Every particle i with charge q_i can subsequently be assumed to be surrounded by a diffuse charge distribution of opposite sign to q_i, with the total charge of the 'cloud' exactly cancelling q_i.

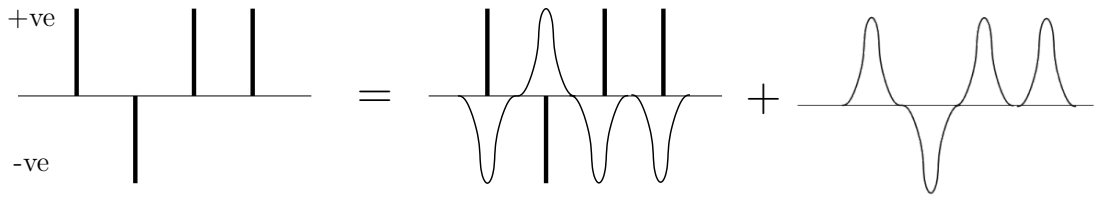


Figure B.1: Representation of the Ewald summation method for a 1-dimensional point charge system, revealing the introduction of Gaussian charge distributions.

The potential experienced at a distance from a point charge is now due to the fraction of q_i not screened by the cloud at that distance. Obviously this function of unscreened charge

rapidly falls to zero as the distance is increased, with the rate being dependent on the functional form of the screening charge distribution. It is usually assumed that this takes the form of a Gaussian distribution of width $\sqrt{2/\alpha}$, (Eq. B.4). This potential can be easily computed by direct summation and is considered the 'real-space' part of the Ewald sum.

$$\rho_{gauss}(r) = -q_i\left(\frac{\alpha}{\pi}\right)^{\frac{3}{2}} exp(-\alpha r^2) \tag{B.4}$$

In order to ensure that the potential due to a set of point charges (and not screened charges) is evaluated, one must now add a term that corrects for the presence of the screening charge via a compensating charge cloud of charge q_i. This can be expressed as a rapidly converging Fourier series and is this regarded as the 'Fourier-space' part of the Ewald sum.

B.2 Fourier-space

In Fourier space one can evaluate the electrostatic potential, $\phi_{screen}(r)$, at a point r_i due to a charge distribution $\rho_{screen}(r)$ (Eq. B.5) due to a periodic sum of Gaussian equations.

$$\rho_{screen}(r) = \sum_{j=1}^{N}\sum_{\mathbf{n}} q_j\left(\frac{\alpha}{\pi}\right)^{\frac{3}{2}} exp[-\alpha|\mathbf{r} - (\mathbf{r}_j + \mathbf{n}L)|^2] \tag{B.5}$$

This can be achieved using Poisson's equation which can be expressed either in its Gaussian form (Eq. B.6) or Fourier form (Eq. B.7).

$$-\nabla^2\phi_{screen}(r) = 4\pi\rho_{screen}(r) \tag{B.6}$$

$$k^2\phi_{screen}(k) = 4\pi\rho_{screen}(k) \tag{B.7}$$

In order to use the latter, one must perform a Fourier transform of the charge density ρ_{screen} (Eq. B.8) and subsequently insert this into Poisson's equation in it's Fourier form (Eq.B.9).

$$\begin{aligned}\rho_{screen}(\mathbf{k}) &= \int_V d\mathbf{r}\ exp(-i\mathbf{k}.\mathbf{r})\rho_{screen}(\mathbf{r}) \\ &= \int_V d\mathbf{r}\ exp(-i\mathbf{k}.\mathbf{r}) \sum_{j=1}^{N}\sum_{\mathbf{n}} q_j \left(\frac{\alpha}{\pi}\right)^{\frac{3}{2}} exp[-\alpha|\mathbf{r}-(\mathbf{r}_j+\mathbf{n}L)|^2] \\ &= \int_{allspace} d\mathbf{r}\ exp(-i\mathbf{k}.\mathbf{r}) \sum_{j=1}^{N} q_j \left(\frac{\alpha}{\pi}\right)^{\frac{3}{2}} exp[-\alpha|\mathbf{r}-\mathbf{r}_j|^2] \\ &= \sum_{j=1}^{N} q_j\ exp(-i\mathbf{k}.\mathbf{r}_j) exp(-k^2/4\alpha) \end{aligned} \tag{B.8}$$

$$\phi_{screen}(k) = \frac{4\pi}{k^2} \sum_{j=1}^{N} q_i\ exp(-i\mathbf{k}.\mathbf{r}_j) exp(-k^2/4\alpha) \tag{B.9}$$

The expression $\phi_{screen}(k)$ is undefined for **k**=0 and is usually assumed to be equal to 0. One must then evaluate the electrostatic potential $\phi_{screen}(r)$ (Eq. B.10) in order to compute the contribution to the potential energy $\mathcal{U}_{screen}$ (Eq. B.11).

$$\begin{aligned}\phi_{screen}(r) &= \frac{1}{V}\sum_{\mathbf{k}\neq 0} \phi_{screen}(\mathbf{k}) exp(i\mathbf{k}.\mathbf{r}) \\ &= \sum_{\mathbf{k}\neq 0}\sum_{j=1}^{N} \frac{4\pi}{Vk^2} q_j\ exp[i\mathbf{k}.(\mathbf{r}-\mathbf{r}_j)] exp(-k^2/4\alpha) \end{aligned} \tag{B.10}$$

$$\begin{aligned}\mathcal{U}_{screen} &= \frac{1}{2}\sum_{i=1}^{N} q_i \phi_{screen}(r_i) \\ &= \frac{1}{2}\sum_{\mathbf{k}\neq 0}\sum_{i,j=1}^{N} \frac{4\pi q_i q_j}{Vk^2} exp[i\mathbf{k}.(\mathbf{r}-\mathbf{r}_j)] exp(-k^2/4\alpha) \end{aligned} \tag{B.11}$$

B.3 Self-Interaction Correction

An important note regarding the computation of $\mathcal{U}_{screen}$ (Eq. B.11) is that this contribution to the potential energy contains a term $(1/2)q_i\phi_{self}(r_i)$. This arises due to the interaction between the point charge q_i and the compensating continuous Gaussian charge cloud of charge q_i. As

previously mentioned it is desired to exclude Coulomb self-interactions and focus only on interactions between a point charge and all other charges in the fundamental box and the periodic images (including its own). Therefore one must correct for this term.

From the over counted charge distribution $\rho_{gauss}(r)$ (Eq. B.12), one can compute the electrostatic potential from the Gaussian form of Poisson's Equation (Eq. B.6) while exploiting the spherical symmetry of the Gaussian cloud (Eq. B.13).

$$\rho_{gauss}(r) = q_i\left(\frac{\alpha}{\pi}\right)^{\frac{3}{2}} exp(-\alpha r^2) \tag{B.12}$$

$$-\frac{1}{r}\frac{\partial^2 r\phi_{gauss}(r)}{\partial r^2} = 4\pi\rho_{gauss}(r) \tag{B.13}$$

A slight rearrangement of this expression (Eq. B.14) allows for a partial integration (Eq. B.15) followed by a second partial integration (Eq. B.16).

$$-\frac{\partial^2 r\phi_{gauss}(r)}{\partial r^2} = 4\pi r\rho_{gauss}(r) \tag{B.14}$$

$$\begin{aligned} -\frac{\partial r\phi_{gauss}(r)}{\partial r} &= \int_{\infty}^{r} dr\ 4\pi r\rho_{gauss}(r) \\ &= -2\pi q_i(\alpha/\pi)^{3/2}\int_{r}^{\infty} dr^2 exp(-\alpha r^2) \\ &= -2q_i(\alpha/\pi)^{1/2} exp(-\alpha r^2) \end{aligned} \tag{B.15}$$

$$r\phi_{gauss}(r) = 2q_i(\alpha/\pi)^{1/2}\int_{0}^{r} dr\ exp(-\alpha r^2) \tag{B.16}$$

Subsequently, by employing the definition of an error function $\text{erf}(x)$ (Eq. B.17) one can evaluate a simplified expression for the electrostatic potential $\phi_{gauss}(r)$ (Eq. B.18).

$$\text{erf}(x) \equiv \frac{2}{\sqrt{\pi}}\int_{0}^{x} exp(-u^2)\ du \tag{B.17}$$

$$\phi_{gauss}(r) = \frac{q_i}{r}\text{erf}(\sqrt{\alpha}r) \tag{B.18}$$

As the point charge q_i is located at the centre of the Gaussian, one need only concern themselves with the value of $\phi_{gauss}(r)$ at $r = 0$ in order to evaluate the ϕ_{self} term (Eq. B.19).

$$\phi_{gauss}(r = 0) = 2q_i(\alpha/\pi)^{1/2} \tag{B.19}$$

This self-interaction potential $\mathcal{U}_{self}$ (Eq. B.20) would subsequently be subtracted from both the 'real-space' and Fourier contributions to the Coulomb energy.

$$\begin{aligned}\mathcal{U}_{self} &= \frac{1}{2}\sum_{i=1}^{N} q_i\phi_{self}(r_i) \\ &= (\alpha/\pi)^{1/2}\sum_{i=1}^{N} q_i^2\end{aligned} \tag{B.20}$$

B.4 Real-Space

The final quantity that one must concern themselves with is the electrostatic energy resultant from the point charges screened by oppositely charged Gaussians. A short-ranged electrostatic potential $\phi_{s.r.}(r)$ can be evaluated (Eq. B.21) which utilises the complementary error function $\text{erfc}(x) \equiv 1-\text{erf}(x)$.

$$\begin{aligned}\phi_{s.r.}(r) &= \frac{q_i}{r} - \frac{q_i}{r}\text{erf}(\sqrt{\alpha}r) \\ &= \frac{q_i}{r}\text{erfc}(\sqrt{\alpha}r)\end{aligned} \tag{B.21}$$

The potential energy due to the screened Coulomb interaction, $\mathcal{U}_{s.r.}$, can be computed (Eq. B.22).

$$\mathcal{U}_{s.r.} = \frac{1}{2}\sum_{i\neq j}^{N} \frac{q_i q_j\ \text{erfc}(\sqrt{\alpha}r_{ij})}{r_{ij}} \tag{B.22}$$

B.5 Total electrostatic potential energy

The total electrostatic contribution to the potential energy $\mathcal{U}_{coulomb}$ is a combination of three terms, $\mathcal{U}_{screen}$, $\mathcal{U}_{self}$ and $\mathcal{U}_{s.r.}$ (Eq. B.23). It should be noted that previously it was stated that one must subtract the $\mathcal{U}_{self}$ term from both the Fourier and 'real-space' expressions. In regards to the 'real-space' term, this is handled via the use of the erfc(x) expression.

$$\begin{aligned}\mathcal{U}_{coulomb} = & \frac{1}{2}\sum_{\mathbf{k}\neq 0}\sum_{i,j=1}^{N}\frac{4\pi q_i q_j}{Vk^2}exp[i\mathbf{k}.(\mathbf{r}-\mathbf{r}_j)]exp(-k^2/4\alpha) \\ & -(\alpha/\pi)^{1/2}\sum_{i=1}^{N}q_i^2 \\ & +\frac{1}{2}\sum_{i\neq j}^{N}\frac{q_i q_j \text{ erfc}(\sqrt{\alpha}r_{ij})}{r_{ij}}\end{aligned} \tag{B.23}$$

C

Publications

P. R. Tulip, C. R. Gregor, R. Z. Troitzsch, G. J. Martyna, E. Cerasoli, G. Tranter, and J. Crain. Conformational Plasticity in an HIV-1 Antibody Epitope. *J. Phys. Chem. B*, 2010, **114** (23).

C. R. Gregor, E. Cerasoli, P. R. Tulip, M. G. Ryadnov, G. J. Martyna, and J. Crain. Autonomous folding in the membrane proximal HIV peptide gp41$_{659-671}$: pH tuneability at micelle interfaces, *Phys. Chem. Chem. Phys.*, 2011, **13**, 127-135.

E. Cerasoli, J. Ravi, C. R. Gregor, R. Hussain, G. Siligardi, G. J. Martyna, J. Crain, and M. G. Ryadnov. Primitive Membrane Mediated Regulation in Free Peptides of HIV-1 gp41: Implications for the Hemifusion Phase. Submitted

C. R. Gregor, E. Cerasoli, J. Schouten, J. Ravi, M. G. Ryadnov, P. Davis, A. Horgan, G. J. Martyna, J. Slootstra, and J. Crain. Antibody recognition of a human chorionic gonadotropin epitope hCGβ_{66-80} depends on local structure retained in the free peptide, *J. Biol. Chem.*, 2011, **286**, 25016-25026.